Bayesian semiparametric Markov renewal mixed models for vocalization syntax

YUTONG WU

Department of Mechanical Engineering, The University of Texas at Austin, TX 78712, USA

ERICH D. JARVIS

Vertebrate Genome Laboratory, Rockefeller University, New York, NY 10065, USA and Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

ABHRA SARKAR*

Department of Statistics and Data Sciences, The University of Texas at Austin, TX 78712, USA
abhra.sarkar@utexas.edu

SUMMARY

Speech and language play an important role in human vocal communication. Studies have shown that vocal disorders can result from genetic factors. In the absence of high-quality data on humans, mouse vocalization experiments in laboratory settings have been proven useful in providing valuable insights into mammalian vocal development, including especially the impact of certain genetic mutations. Such data sets usually consist of categorical syllable sequences along with continuous intersyllable interval (ISI) times for mice of different genotypes vocalizing under different contexts. ISIs are of particular importance as increased ISIs can be an indication of possible vocal impairment. Statistical methods for properly analyzing ISIs along with the transition probabilities have however been lacking. In this article, we propose a class of novel Markov renewal mixed models that capture the stochastic dynamics of both state transitions and ISI lengths. Specifically, we model the transition dynamics and the ISIs using Dirichlet and gamma mixtures, respectively, allowing the mixture probabilities in both cases to vary flexibly with fixed covariate effects as well as random individual-specific effects. We apply our model to analyze the impact of a mutation in the Foxp2 gene on mouse vocal behavior. We find that genotypes and social contexts significantly affect the length of ISIs but, compared to previous analyses, the influences of genotype and social context on the syllable transition dynamics are weaker.

Keywords: Clustering; Dirichlet mixtures; Gamma mixtures; Markov renewal processes; Mixed effects models; Mouse vocalization experiments.

1. INTRODUCTION

Spoken language plays a crucial role in almost every aspect of human life as we use it to share information, communicate ideas, and express emotions. However, our vocal behaviors might be restrained

---

*To whom correspondence should be addressed.

by a wide variety of impairments, some of which are highly inheritable. According to the National Institute on Deafness and Other Communication Disorders, the prevalence of speech and sound disorder among young children is 8–9% and the majority of such disorders have no known cause (NIH-NIDCD Report, 2020). As many speech-related disorders are inheritable (Vargha-Khadem and others, 1998), it is thus important to study the genetic and evolutionary development of human vocal communication to identify and remedy vocal disorders.

Since data on human vocalization disorders are not easily available, neither can humans be studied under experimentally induced impairing conditions, neuroscientists have turned to studying mouse vocalization systems to gain insights into human vocal communication processes (Jarvis, 2019). Although unlike speech, the mouse vocalization is mostly innate (Arriaga and Jarvis, 2013; Mooney, 2020), it is a particularly attractive model to study for several reasons. Adult mice "sing" ultrasonic vocalizations (USVs) to communicate with each other. Being mammals, they are also physiologically and genetically similar to us humans. The patterns of USVs may be influenced by the mouse genotype or environmental factors such as stimulating social contexts (Chabout and others, 2015). Mouse vocalization data sets thus typically comprise songs sung by mice from different genotypes under various social contexts. Systematic differences in the syllable dynamics across various genotypes and social conditions can provide insights into their roles on vocal abilities and behavior (Chabout and others, 2016).

Our research is motivated by the need for sophisticated statistical methods for analyzing mouse vocalization syntax generated in laboratory experiments that are conducted to understand the effects of certain genetic mutations and social contexts on mouse vocal behavior. We introduce a novel class of Bayesian mixed models for analyzing categorical sequences with continuous interstate interval times under the influence of multiple exogenous factors. In particular, the values of the exogenous factors remain fixed throughout the sequence and contributes a fixed group effect. Each sequence is also associated with an individual that exhibits a random individual effect. We are interested in the inference of the stochastic dynamics of the sequences, specifically, the transition dynamics of the discrete states as well as the distribution of the continuous interstate interval times. Statistical methods for analyzing the syllable dynamics have previously been developed by Holy and Guo (2005) and Chabout and others (2015, 2016). Sarkar and others (2018) developed a flexible Bayesian mixed effects Markov model for vocalization syntax incorporating exogenous influences of covariates as well as random heterogeneity of the sampled mice. While these methods looked in detail into the systematic differences in the syllable dynamics, they did not properly analyze the intersyllable intervals (ISIs) which can be an additional important indicator of vocal deficits as impaired mice will tend to remain silent with longer ISIs. To accommodate this effect, the aforementioned methods discretized large ISIs into one or multiple special "silent" syllables and then treated the songs as Markov sequences with this appended vocabulary. This practice, however, mixes the influences of covariates on transition probabilities and ISIs which could result in less accurate scientific conclusions. The previous methods largely ignore the differences in the distributions of the ISIs from different mice under various combinations of the covariate values and may miss out important evidence that can be deduced by properly modeling the ISIs.

Here, we develop an approach to address these concerns by appropriately analyzing the syllable transition dynamics using a slightly modified version of the mixed Markov model of Sarkar and others (2018) while separately modeling the distribution of ISIs using a novel flexible mixed model of gamma mixtures, thereby providing inference for both syllable transitions dynamics and their ISIs. The method accommodates fixed covariate effects as well as random individual effects in both the syllable transition dynamics and the ISI distributions. A hierarchical cluster inducing mechanism for the levels of the covariates allows straightforward, formal tests of their significance. We design an efficient Markov chain Monte Carlo sampler for fitting our model. We demonstrate the performance

of our model by analyzing a data set where the mice are either wild-type or carry a mutation on the Foxp2 gene implicated in causing vocal impairment in humans. Previous analyses of this data set by Chabout *and others* (2015, 2016) and Sarkar *and others* (2018), with large ISIs treated as an artificial syllable, have shown significant differences in the syllable dynamics between different genotypes and social contexts. When reanalyzed using our proposed approach, the results suggest that genotype and context strongly impact the ISIs, but their influences on the syllable transition dynamics are weaker than what previous analyses had inferred.

Our proposed model is a type of Markov renewal process (MRP) where the state transitions evolve as a Markov chain while the state durations follow a transition density function that depends on both the previous and the current state. MRPs were originally introduced by Pyke (1961). MRPs and their variations, including semi-Markov models (Levy, 1954; Smith, 1955), have found success in a variety of applications such as in modeling clinical trials (Weiss and Zelen, 1965), sleeping patterns (Yang and Hursch, 1973), HIV disease occurrences (Foucher *and others*, 2005), etc. Bayesian methods for MRPs have also been developed. Phelan (1990) designed an MRP where the prior consisted of a family of Dirichlet distributions for transition matrices and a Beta family of Levy processes for state duration times. Muliere *and others* (2003) and Bulla and Muliere (2007) developed Bayesian nonparametric reinforced MRPs. Bayesian MRPs with Weibull distributed interoccurrence times have been developed for seismic data in Alvarez (2005); Epifani *and others* (2014). Unlike most classical MRPs that focus on modeling a single sequence, we jointly model a collection of sequences, each one associated with an individual as well as a set of time-invariant external covariates, accommodating fixed effects of the covariates as well as random effects of the subjects for both the transition probability matrices and the distribution of the ISIs. We also allow the selection of important covariates for both the state transition dynamics and the ISI distribution via probabilistic partitioning of the covariate levels. These are in contrast to existing methods on classical MRPs where typically only a single sequence is modeled and the distribution of state duration depends only on the current state.

Besides being directly useful for analyzing mouse vocalization data sets, the methodology proposed in the article can be used for a great variety of applications where MRPs are useful such as the examples cited above. Our proposed model for the ISIs may also be of independent statistical interest in developing mixture models for continuous variables with mixed covariate and individual effects which, to our knowledge, have not been explored much in the literature.

The article is organized as follows. In Section 2, we provide details of the Foxp2 data set and its scientific background and review some previous statistical methods. In Section 3, we present our novel Bayesian Markov renewal mixed model. In Section 4, we briefly outline our Markov chain Monte Carlo (MCMC) algorithm to sample from the posterior. We illustrate the results of our method applied to the Foxp2 data set in Section 5. Section 6 concludes with a discussion.

2. Data set and preliminaries

The Foxp2 (foxhead-box P2) gene is a transcription factor that regulates other genes (Chabout *and others*, 2016). It is found in mice and in similar forms in humans (Fisher and Scharff, 2009), and its mutation has been implicated to cause speech and language deficits in adults (Lai *and others*, 2001). Previously, Fujita *and others* (2008) compared the vocalizations for wild-type, heterozygous Foxp2 and homozygous Foxp2 mice and showed that both heterozygous and homozygous Foxp2 mice have vocal impairment to some extent. Castellucci *and others* (2016) showed that mice with a Foxp2 mutation vocalize less and produce shorter syllable sequences. Gaub *and others* (2016) discovered that compared to wild-type, mice with Foxp2 mutations displayed quantitative differences in USVs. Other than genotypes, social contexts can also influence mouse vocalization. Chabout *and others* (2012) showed that the amount of USVs emitted by male mice is positively correlated with
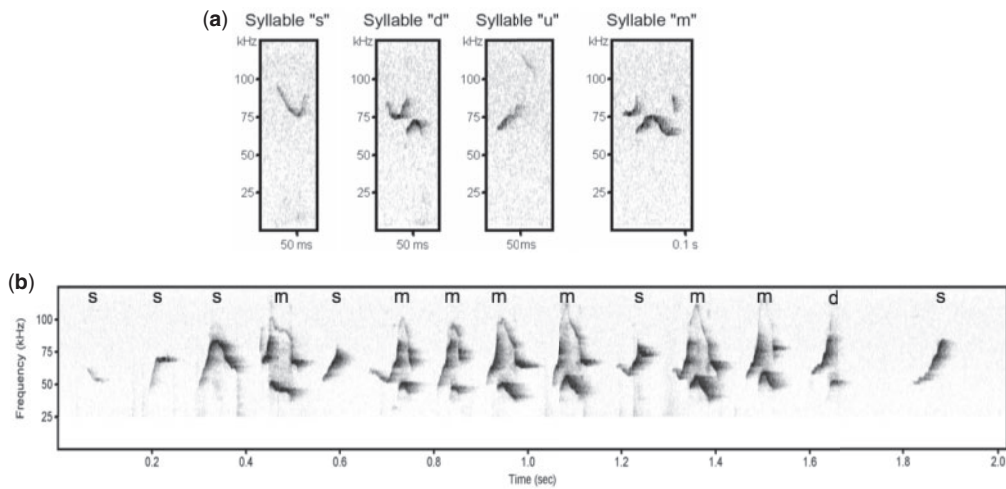
Fig. 1. Spectral diagrams of mouse vocalizations, reproduced from Chabout *and others* (2016) with permission. (a) Syllable types. (b) Part of a song produced by a wild-type male mouse under the urine context (U).

the scale of their social interactions. Gaub *and others* (2016) studied the USVs of adult mice with increasing stimulus intensity including water and female urine.

The overall properties of the song syllables provide some information. The sequencing rules of the syllables provide additional information about song complexity. Research has thus also been conducted on the vocabulary and structure of the songs, referred to as "syntax" (Holy and Guo, 2005; Moles *and others*, 2007; Scattoni *and others*, 2011; Musolf *and others*, 2015; Gaub *and others*, 2016). Adopting the vocabulary in Chabout *and others* (2012, 2015, 2016), the mouse vocalization syllables can be grouped into four categories based on their spectral features (Figure 1a) as: "s": simple syllables without any pitch jumps; "u": complex syllables with a single upward pitch jump; "d": complex syllables with a single downward pitch jump; and "m": more complex syllables with a series of multiple pitch jumps. Songs are made of a sequential arrangements of these syllables (Figure 1b) with ISIs varying mostly between 0 and 250 ms (Table 1a, Figure S1 in Appendix A of the Supplementary material available at *Biostatistics* online).

The Foxp2 data set that we analyze here collects the songs produced by mice of wild-type (W) and mice that have a mutation on the Foxp2 gene (F) under three different social contexts—fresh female urine on a cotton tip placed inside the male's cage (U), awake and behaving adult female placed inside the cage (L), and one anesthetized female placed on the lid of the cage (A). The data set has 70 818 rows, including 49 songs sung by 18 mice, 10 with the Foxp2 mutation, and 8 wild-types. Five mice sang 2 songs each (with id 1, 2, 4, 10, and 11), and the rest 13 mice sang 3 songs each. Each mouse sang the songs under different contexts. Therefore, the combination of covariates and individuals associated with each song is actually unique. The distribution of songs across different combinations of covariates is presented in Table 1b (left). There is no missing or censored data. The empirical distribution of syllable transitions is displayed in Table 1b (right). The number of transition types stratified by genotypes and social contexts can be found in the Appendix A of the Supplementary material available at *Biostatistics* online.

Since the complexity of the four syllables vary, it is reasonable to assume that mice with vocal impairments will produce songs with fewer transitions to difficult syllables such as m. It is also

Table 1. *Description of the Foxp2 data set*
(a) Part of the data set associated with the anesthetized female context (*A*).

Mouse ID	Genotype	Syllable	ISI (in s)
1	*F*	s	0.082
1	*F*	s	0.017
1	*F*	m	0.114
⋮	⋮	⋮	⋮
2	*W*	s	1.546
2	*W*	d	0.712
2	*W*	s	0.549

(b) Empirical distributions of the Foxp2 data set: song distribution across different combinations of covariates (left) and number of transitions for all pairs of syllables (right).

Distribution of songs

Genotype	Context	No. of songs
F	*U*	8
F	*L*	10
F	*A*	10
W	*U*	6
W	*L*	8
W	*A*	7

Distribution of syllable transitions

		current syllable			
		d	m	s	u
Preceding syllable	d	2780	964	5980	268
	m	987	983	3187	257
	s	5920	3213	42 138	1742
	u	305	257	1705	132

(c) The empirical and posterior (in parentheses) means and standard deviations of the ISIs, grouped by covariate levels.

	Mean	s.d.		Mean	s.d.
F	0.228 (0.227)	0.433 (0.417)	*U*	0.257 (0.237)	0.504 (0.424)
W	0.161 (0.163)	0.333 (0.320)	*L*	0.166 (0.168)	0.319 (0.335)
			A	0.211 (0.218)	0.426 (0.399)

	Mean	s.d.		Mean	s.d.
(d,d)	0.134 (0.127)	0.236 (0.241)	(s,d)	0.168 (0.154)	0.341 (0.303)
(d,m)	0.119 (0.126)	0.261 (0.239)	(s,m)	0.131 (0.129)	0.253 (0.246)
(d,s)	0.180 (0.181)	0.355 (0.351)	(s,s)	0.217 (0.221)	0.423 (0.407)
(d,u)	0.172 (0.144)	0.343 (0.300)	(s,u)	0.167 (0.151)	0.338 (0.294)
(m,d)	0.118 (0.138)	0.234 (0.267)	(u,d)	0.130 (0.157)	0.290 (0.319)
(m,m)	0.120 (0.129)	0.231 (0.245)	(u,m)	0.126 (0.130)	0.210 (0.246)
(m,s)	0.145 (0.144)	0.295 (0.301)	(u,s)	0.165 (0.144)	0.373 (0.303)
(m,u)	0.120 (0.155)	0.228 (0.305)	(u,u)	0.139 (0.174)	0.294 (0.341)

expected that mice with vocal impairments tend to remain more silent with longer ISIs. Table 1c shows the empirical distribution of the mean and standard deviation of ISIs grouped by each covariate. We see that mice with the Foxp2 mutation are likely to have longer ISI than that of wild-types, with a 42% increase in the empirical distribution of ISI mean.

The development of sophisticated statistical methods for mouse vocalization syntax started with Holy and Guo (2005) who analyzed the songs using a Markov model. Chabout *and others* (2015, 2016) developed statistical tests for accessing global and local syntax differences across genotypes and social contexts. Sarkar *and others* (2018) developed a mixed Markov model for the transitions of four syllables and an extra artificial syllable for large ISIs. In the latter three works, each ISI of

length greater than 250 ms was treated as a silent state ("x"). Inference was then performed treating the resulting sequences as Markov chains with five states {d, m, s, u, x}. Though this was done to achieve significant analytical convenience, the resulting stochastic dynamics also got dominated by transitions to x. Moreover, the distribution of the ISIs varied greatly between different experimental conditions and subjects (Figure S1 in Appendix A of the Supplementary material available at *Biostatistics* online). Ignoring the ISIs with lengths shorter than 250 ms, as well as treating longer ISIs as blocks of silent syllables, resulted in loss of important information in addition to diluting the transition dynamics among the original syllables. In order to obtain a more accurate inference of the data set, it is important to treat the ISIs differently from the four original syllables and model them properly as a continuous variable.

Our proposed approach addresses these concerns. We model the ISIs separately instead of treating it ad hocly as the "silence" syllable. In this way, the ISIs can be used as evidence for vocal impairment aside from the transitions of the four syllables. Moreover, we allow both the transition probabilities and the mixture probabilities for the ISIs to be governed by a convex combination of population-level fixed effects and individual-level random effects.

3. MARKOV RENEWAL MIXED MODELS

Consider a sequence s of T_s syllables. $y_{s,t}$ denotes the syllable at time t for sequence s and is one of $\mathcal{Y} = \{d, m, s, u\} = \{1, 2, 3, 4\}$. The collection of syllables is denoted by $\{y_{s,t}\}_{s=1,t=1}^{s_0,T_s}$, where s_0 is the total number of sequences. Within a sequence s, we have $T_s - 1$ ISI times, denoted by $\{\tau_{s,t}\}_{s=1,t=2}^{s_0,T_s}$, where $\tau_{s,t}$ represents the interval time between the $(t-1)$th and tth syllable for sequence s. Each sequence s is generated under two exogenous factors—genotype $x_{s,1} \in \mathcal{X}_1 = \{F, W\} = \{1, 2\}$, and social context $x_{s,2} \in \mathcal{X}_2 = \{U, L, A\} = \{1, 2, 3\}$, as described in Section 2. With some abuse, we use the same notation to denote the variables as well as their specific values, greatly simplifying the exposition.

.

Notations for the Foxp2 data set	
$\mathcal{X}_1$	Set of genotypes, $\{F, W\}$
$\mathcal{X}_2$	Set of social contexts, $\{U, L, A\}$
$\mathcal{Y}$	Set of four sound syllables, $\{d, m, s, u\}$
$y_{s,t}$	Sound syllable at time t for sequence s, $y_{s,t} \in \mathcal{Y}$
$\tau_{s,t}$	ISI between $(t-1)$th and tth syllable of sequence s
$x_{s,1}$	Covariate 1, genotype, $x_{s,1} \in \mathcal{X}_1$
$x_{s,2}$	Covariate 2, social context, $x_{s,2} \in \mathcal{X}_2$
$(y_{s,t-1}, y_{s,t})$	Covariate 3, the preceding and the current syllable, $y_{s,t-1} \in \mathcal{Y}, y_{s,t} \in \mathcal{Y}$
Notations for the transition dynamics of the syllables	
$z_{\text{trans},j,x_{s,j}}$	Cluster label of $x_{s,j}$
$\mathcal{C}_{\text{trans}}^{(j)}$	Partition of $\mathcal{X}_j$
$k_{\text{trans},j}$	Number of clusters of partition $\mathcal{C}_{\text{trans}}^{(j)}$
$\boldsymbol{\lambda}_{\text{trans},x_1,x_2}$	Transition probability vector for covariate x_1 and x_2
$\boldsymbol{\lambda}_{\text{trans},h_1,h_2}$	Transition probability vector for cluster h_1 and h_2
$\boldsymbol{\lambda}_{\text{trans}}^{(i)}$	Transition probability vector for mouse i
$\boldsymbol{\pi}_{\text{trans},0}^{(i)}$	State-specific probability of fixed effect of mouse i
$\boldsymbol{\pi}_{\text{trans},1}^{(i)}$	State-specific probability of random effect of mouse i

.

Notations for the distributions of the ISIs

|	
$z_{isi,s,t}$	Component label of gamma mixtures
$z_{isi,r,x_{s,r}}$	Cluster label of $x_{s,r}$
$z_{isi,r,(y_{s,t-1},y_{s,t})}$	Cluster label of $(y_{s,t-1}, y_{s,t})$
$\mathcal{C}_{isi}^{(r)}$	Partition of $\mathcal{X}_r/\mathcal{Y}$
$k_{isi,r}$	Number of clusters of partition $\mathcal{C}_{isi}^{(r)}$
$\lambda_{isi,x_1,x_2,(y_{t-1},y_t)}$	Mixture probability vector for covariate x_1, x_2 and syllables (y_{t-1}, y_t)
$\lambda_{isi,g_1,g_2,g_3}$	Mixture probability vector for cluster g_1, g_2 and g_3
$\lambda_{isi}^{(i)}$	Mixture probability vector for mouse i
$\pi_{isi,0}^{(i)}$	Component-specific probability of fixed effect of mouse i
$\pi_{isi,1}^{(i)}$	Component-specific probability of random effect of mouse i
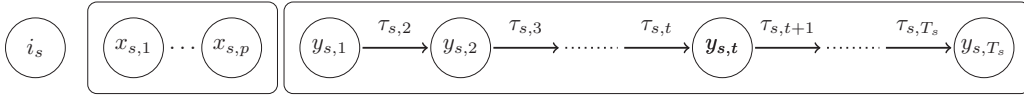
Fig. 2. Graphical model showing the data structures: $y_{s,t}$ denotes the observed state at the tth time location in the sth sequence and $\tau_{s,t}$ denotes the observed ISIs between the states $y_{s,t-1}$ and $y_{s,t}$; each sequence s is also associated with an individual i_s and a set of exogenous time-invariant covariates $x_{s,1}, \ldots, x_{s,p}$. The Markov renewal mixed model considered in this article analyzes the state transitions $y_{s,t}$ and the ISI lengths $\tau_{s,t}$, accommodating fixed effects of the covariates $x_{s,1}, \ldots, x_{s,p}$ and random effects of the individuals i_s.

We recall that for each sequence s, $(\mathbf{y}, \boldsymbol{\tau})$ is a MRP when we have

$$p(y_{s,t}, \tau_{s,t} \mid y_{s,t-1}, \tau_{s,t-1}, \ldots, y_{s,1}, \tau_{s,1}, \boldsymbol{\theta}_{\text{trans}}, \boldsymbol{\theta}_{isi}) = p(y_{s,t}, \tau_{s,t} \mid y_{s,t-1}, \boldsymbol{\theta}_{\text{trans}}, \boldsymbol{\theta}_{isi}),$$

where $\boldsymbol{\theta}_{\text{trans}}$ and $\boldsymbol{\theta}_{isi}$ are the parameters specifying the transition and ISI distributions, respectively. We further assume that (i) the ISI lengths $\tau_{s,t}$'s depend on both the preceding syllable $y_{s,t-1}$ and the current syllable $y_{s,t}$ (Figure 2) and (ii) the state transitions and the ISI densities do not share any parameters, that is, $\boldsymbol{\theta}_{\text{trans}} \cap \boldsymbol{\theta}_{isi} = \phi$. Then, we can further write

$$p(y_{s,t}, \tau_{s,t} \mid y_{s,t-1}, \boldsymbol{\theta}_{\text{trans}}, \boldsymbol{\theta}_{isi}) = p(y_{s,t} \mid y_{s,t-1}, \boldsymbol{\theta}_{\text{trans}}) \cdot p(\tau_{s,t} \mid y_{s,t-1}, y_{s,t}, \boldsymbol{\theta}_{isi}).$$

With $p_0(\cdot)$ denoting the priors, the joint posterior of $(\boldsymbol{\theta}_{\text{trans}}, \boldsymbol{\theta}_{isi})$ then also factorizes as

$$p(\boldsymbol{\theta}_{\text{trans}}, \boldsymbol{\theta}_{isi} \mid \mathbf{y}, \boldsymbol{\tau}) \propto p(\mathbf{y} \mid \boldsymbol{\theta}_{\text{trans}}) p(\boldsymbol{\tau} \mid \boldsymbol{\theta}_{isi}, \mathbf{y}) p_0(\boldsymbol{\theta}_{\text{trans}}) p_0(\boldsymbol{\theta}_{isi}) \propto p(\boldsymbol{\theta}_{\text{trans}} \mid \mathbf{y}) p(\boldsymbol{\theta}_{isi} \mid \mathbf{y}, \boldsymbol{\tau}),$$

Going forward, we thus discuss the models for the syllables and the ISIs separately.

3.1. *Model for syllable transitions*

We use the Bayesian mixed effects Markov model of Sarkar *and others* (2018) for syllable transitions. We describe the model here to keep this article relatively self-contained.

We begin with specifying the transition dynamics as a mixed Markov model as

$$\Pr(y_{s,t} = y_t \mid i_s = i, x_{s,1} = x_1, x_{s,2} = x_2, y_{s,t-1} = y_{t-1}) = P_{\text{trans},x_1,x_2}^{(i)}(y_t \mid y_{t-1}),$$

$$P_{\text{trans},x_1,x_2}^{(i)}(y_t \mid y_{t-1}) = \pi_{\text{trans},0}^{(i)}(y_{t-1})\lambda_{\text{trans},x_1,x_2}^{(i)}(y_t \mid y_{t-1}) + \pi_{\text{trans},1}^{(i)}(y_{t-1})\lambda_{\text{trans}}^{(i)}(y_t \mid y_{t-1}).$$

(3.1)

The mixed effects transition probabilities $P_{\text{trans},x_1,x_2}^{(i)}(y_t \mid y_{t-1})$'s are modeled here as a flexible convex mixture of a baseline fixed effect component $\boldsymbol{\lambda}_{\text{trans},x_1,x_2}(\cdot \mid y_{t-1})$ for the exogenous covariates, namely genotype and context, and a random effect component $\boldsymbol{\lambda}_{\text{trans}}^{(i)}(\cdot \mid y_{t-1})$ for the mouse. The weights $\pi_{\text{trans},0}^{(i)}$ and $\pi_{\text{trans},1}^{(i)} = 1 - \pi_{\text{trans},0}^{(i)}$ of the two effects are also allowed to be mouse specific. In Sarkar *and others* (2018), the coefficient for the convex combination did not vary between the individuals but was fixed at $\pi_{\text{trans},1}(y_{t-1})$. The population-level model obtained after integrating out the $\boldsymbol{\lambda}_{\text{trans}}^{(i)}(y_t \mid y_{t-1})$'s was shown to be able to characterize all possible cases of predictor dependent transition probabilities, including accommodating all order interactions between them, and the individual-level model could also accommodate deviations from it very flexibly. Our specification here retains the nonparametric nature of the population-level model but by allowing the coefficients $\pi_{\text{trans},1}^{(i)}(y_{t-1})$ to be mouse specific, which accommodates more flexibility in characterizing individual heterogeneity.

For the fixed effects of each covariate j, we try to further identify its levels that have a similar effect on the song dynamics. This is done by creating a probabilistic partition $\mathcal{C}_{\text{trans}}^{(j)} = \{\mathcal{C}_{\text{trans},h_j}^{(j)}\}_{h_j=1}^{k_{\text{trans},j}}$ of its levels. Given partitions $\mathcal{C}_{\text{trans}}^{(1)}$ and $\mathcal{C}_{\text{trans}}^{(2)}$, songs with covariates in the same clusters, say $\mathcal{C}_{\text{trans},h_1}^{(1)}$ and $\mathcal{C}_{\text{trans},h_2}^{(2)}$, then share the same baseline transition probability $\boldsymbol{\lambda}_{\text{trans},h_1,h_2}(\cdot \mid y_{t-1})$. The specification of the probabilistic partition models is facilitated by introducing latent cluster allocation variables $\{z_{\text{trans},j,\ell}\}_{j=1,\ell=1}^{2,d_j}$, with $z_{\text{trans},j,\ell}$ indicating the cluster label for the ℓth level of the jth covariate. Two levels $\ell_1, \ell_2 \in \mathcal{X}_j = \{1, \ldots, d_j\}$ will be clustered together if and only if $z_{\text{trans},j,\ell_1} = z_{\text{trans},j,\ell_2}$. For example, $z_{\text{trans},j=2,\ell=2} = z_{\text{trans},j=2,\ell=3} = 1$ means that songs produced under contexts L ($j = 2, \ell = 2$) and A ($j = 2, \ell = 3$) belong to cluster $\mathcal{C}_{\text{trans},h_2=1}^{(j=2)}$ of $\mathcal{C}_{\text{trans}}^{(j=2)}$, etc. Importantly, when the levels of a covariate j are all clustered together, that is, $k_{\text{trans},j} = 1$, the transition probabilities do not vary with the levels of covariate j. The covariate j thus has no effect on the transition dynamics when $k_{\text{trans},j} = 1$, allowing us to easily and formally test its significance based on the posterior probability of the event $k_{\text{trans},j} = 1$.

The final transition probability of syllables in a song s produced by a mouse $i_s = i$ with genotype $x_{s,1}$ in cluster h_1 and context $x_{s,2}$ in cluster h_2 is given by

$$\mathbf{P}_{\text{trans},h_1,h_2}^{(i)}(\cdot \mid y_{t-1}) = \pi_{\text{trans},0}^{(i)}(y_{t-1})\boldsymbol{\lambda}_{\text{trans},h_1,h_2}(\cdot \mid y_{t-1}) + \pi_{\text{trans},1}^{(i)}(y_{t-1})\boldsymbol{\lambda}_{\text{trans}}^{(i)}(\cdot \mid y_{t-1}).$$

We assign conditionally conjugate Dirichlet priors to the fixed effect components

$$\boldsymbol{\lambda}_{\text{trans},h_1,h_2}(\cdot \mid y_{t-1}) \sim \text{Dir}\left\{\alpha_{\text{trans},0}\lambda_{\text{trans},0}(1 \mid y_{t-1}), \ldots, \alpha_{\text{trans},0}\lambda_{\text{trans},0}(4 \mid y_{t-1})\right\}.$$

For the random effect distribution, for any $y_{t-1} \in \mathcal{Y}$, we let

$$\boldsymbol{\lambda}_{\text{trans}}^{(i)}(\cdot \mid y_{t-1}) \sim \text{Dir}\left\{\alpha_{\text{trans}}^{(0)}\lambda_{\text{trans},0}(1 \mid y_{t-1}), \ldots, \alpha_{\text{trans}}^{(0)}\lambda_{\text{trans},0}(4 \mid y_{t-1})\right\}.$$

We assign a Beta prior to the mouse-specific coefficient for any $y_{t-1} \in \mathcal{Y}$ as $\pi_{\text{trans},0}^{(i)}(y_{t-1}) \sim$ Beta$(a_{\text{trans},0}, a_{\text{trans},1})$. Centering both $\boldsymbol{\lambda}_{\text{trans},h_1,h_2}$ and $\boldsymbol{\lambda}_{\text{trans}}^{(i)}$ around the same mean vector $\boldsymbol{\lambda}_{\text{trans},0}$ facilitates posterior computation. The random effects $\boldsymbol{\lambda}_{\text{trans}}^{(i)}$ and $\boldsymbol{\pi}_{\text{trans},0}^{(i)}$ can be easily integrated out to obtain a closed-form expression for population-level probabilities as

$$\mathbf{P}_{\text{trans},h_1,h_2}(\cdot \mid y_{t-1}) = \pi_{\text{trans},0}\boldsymbol{\lambda}_{\text{trans},h_1,h_2}(\cdot \mid y_{t-1}) + \pi_{\text{trans},1}\boldsymbol{\lambda}_{\text{trans},0}(\cdot \mid y_{t-1}),$$

where $\pi_{\text{trans},0} = \frac{a_{\text{trans},0}}{a_{\text{trans},0}+a_{\text{trans},1}}$, $\pi_{\text{trans},1} = \frac{a_{\text{trans},1}}{a_{\text{trans},0}+a_{\text{trans},1}}$. Some states in $\mathcal{Y}$ are naturally preferred regardless of the values of the covariates. To capture this, we let $\boldsymbol{\lambda}_{\text{trans},0}$ center around a global $\boldsymbol{\lambda}_{\text{trans},00}$. Lastly, the hyperparameters $\alpha_{\text{trans}}^{(0)}$ and $\alpha_{\text{trans},0}$ are given gamma hyperpriors.

The complete Bayesian hierarchical model for the transitions can be summarized as follows.

$$
(y_{s,t} \mid y_{s,t-1} = y_{t-1}, i_s = i, z_{\text{trans},1,x_{s,1}} = h_1, z_{\text{trans},2,x_{s,2}} = h_2) \sim
$$
$$
\text{Mult}\left\{P_{\text{trans},h_1,h_2}^{(i)}(1 \mid y_{t-1}), \ldots, P_{\text{trans},h_1,h_2}^{(i)}(4 \mid y_{t-1})\right\},
$$
$$
\mathbf{P}_{\text{trans},h_1,h_2}^{(i)}(\cdot \mid y_{t-1}) = \pi_{\text{trans},0}^{(i)}(y_{t-1})\boldsymbol{\lambda}_{\text{trans},h_1,h_2}(\cdot \mid y_{t-1}) + \pi_{\text{trans},1}^{(i)}(y_{t-1})\boldsymbol{\lambda}_{\text{trans}}^{(i)}(\cdot \mid y_{t-1}),
$$
$$
z_{\text{trans},j,\ell} \sim \text{Mult}\left\{\mu_{\text{trans},j}(1), \ldots, \mu_{\text{trans},j}(d_j)\right\}, \quad \boldsymbol{\mu}_{\text{trans},j} \sim \text{Dir}(\alpha_{\text{trans},j}, \ldots, \alpha_{\text{trans},j}),
$$
$$
\boldsymbol{\lambda}_{\text{trans}}^{(i)}(\cdot \mid y_{t-1}) \sim \text{Dir}\left\{\alpha_{\text{trans}}^{(0)}\lambda_{\text{trans},0}(1 \mid y_{t-1}), \ldots, \alpha_{\text{trans}}^{(0)}\lambda_{\text{trans},0}(4 \mid y_{t-1})\right\}, \tag{3.2}
$$
$$
\boldsymbol{\lambda}_{\text{trans},h_1,h_2}(\cdot \mid y_{t-1}) \sim \text{Dir}\left\{\alpha_{\text{trans},0}\lambda_{\text{trans},0}(1 \mid y_{t-1}), \ldots, \alpha_{\text{trans},0}\lambda_{\text{trans},0}(4 \mid y_{t-1})\right\},
$$
$$
\boldsymbol{\lambda}_{\text{trans},0}(\cdot \mid y_{t-1}) \sim \text{Dir}\left\{\alpha_{\text{trans},00}\lambda_{\text{trans},00}(1), \ldots, \alpha_{\text{trans},00}\lambda_{\text{trans},00}(4)\right\},
$$
$$
\pi_{\text{trans},0}^{(i)}(y_{t-1}) \sim \text{Beta}(a_{\text{trans},0}, a_{\text{trans},1}), \quad \alpha_{\text{trans},0} \sim \text{Ga}(a_{\text{trans},0}, b_{\text{trans},0}), \quad \alpha_{\text{trans}}^{(0)} \sim \text{Ga}(a_{\text{trans}}^{(0)}, b_{\text{trans}}^{(0)}).
$$

3.2. *Model for ISIs*

For the ISI times $\{\tau_{s,t}\}_{s=1,t=2}^{s_0,T_s}$, we associate each ISI from song s with three covariates: two exogenous, namely genotype $x_{s,1} \in \mathcal{X}_1 = \{1, 2\}$ and social context $x_{s,2} \in \mathcal{X}_2 = \{1, 2, 3\}$, and one local, namely the pair of the preceding and the current syllable $(y_{s,t-1}, y_{s,t}) \in \mathcal{Y} \times \mathcal{Y} = \{1, 2, 3, 4\} \times \{1, 2, 3, 4\}$. Given the values of these covariates, we model the log-transformed ISI times $\widetilde{\tau}_{s,t} = \log(\tau_{s,t} + 1)$ of the mouse $i_s = i$ using mixtures of gamma kernels as

$$
f(\widetilde{\tau}_{s,t} \mid i_s = i, x_{s,1} = x_1, x_{s,2} = x_2, (y_{s,t-1}, y_{s,t}) = (y_{t-1}, y_t))
$$
$$
= \sum_{k=1}^{K} P_{isi}^{(i)}(k \mid x_1, x_2, (y_{t-1}, y_t))\text{Ga}(\widetilde{\tau}_{s,t} \mid \alpha_k, \beta_k), \tag{3.3}
$$

where $\text{Ga}(\cdot \mid \alpha_k, \beta_k)$ denotes a gamma mixture kernel with shape α_k and rate β_k and K is the total number of mixture components. $P_{isi}^{(i)}(k \mid x_1, x_2, (y_{t-1}, y_t))$'s are mixed effects mixture probabilities that vary with the associated covariate values and are also specific to the subject. Introducing latent variables $\{z_{isi,s,t}\}$ indicating the index of the mixture component, we can write

$$
f(\widetilde{\tau}_{s,t} \mid z_{isi,s,t} = k) \sim \text{Ga}(\widetilde{\tau}_{s,t} \mid \alpha_k, \beta_k),
$$
$$
\Pr(z_{isi,s,t} = k \mid i_s = i, x_{s,1} = x_1, x_{s,2} = x_2, (y_{s,t-1}, y_{s,t}) = (y_{t-1}, y_t)) = P_{isi}^{(i)}(k \mid x_1, x_2, (y_{t-1}, y_t)). \tag{3.4}
$$

Model (3.4) is structurally similar to model (3.1) except that we are now modeling the distribution of a *latent* categorical variable $z_{isi,s,t}$ as opposed to the *observed* categorical variable $y_{s,t}$. The number of components K in (3.3) is thus also unknown and needs to be inferred from the data, bringing in significant additional challenges. Nevertheless, we can use similar strategies to model the mixed effects mixture probabilities $P_{isi}^{(i)}(k \mid x_1, x_2, (y_{t-1}, y_t))$ as

$$
\mathbf{P}_{isi,x_1,x_2,(y_{t-1},y_t)}^{(i)}(\cdot) = \pi_{isi,0}^{(i)}(\cdot)\boldsymbol{\lambda}_{isi,x_1,x_2,(y_{t-1},y_t)}(\cdot) + \pi_{isi,1}^{(i)}(\cdot)\boldsymbol{\lambda}_{isi}^{(i)}(\cdot),
$$

where $\boldsymbol{\lambda}_{isi,x_1,x_2,(y_{t-1},y_t)}(\cdot)$ is the fixed effect component for the associated covariates, namely genotype, context and the preceding–current syllable pair, and $\boldsymbol{\lambda}_{isi}^{(i)}(\cdot)$ is the random effect component for the

mouse. The weights $\pi_{isi,0}^{(i)}$ and $\pi_{isi,1}^{(i)} = 1 - \pi_{isi,0}^{(i)}$ of the two effects are also allowed to be mouse specific, as before.

To assess the significance of each covariate r, we induce a clustering $\mathcal{C}_{isi}^{(r)} = \{\mathcal{C}_{isi,gr}^{(r)}\}_{gr=1}^{k_{isi,r}}$ of its levels so that ISIs with associated covariates in the same clusters, say g_1, g_2, and g_3, share the same fixed effect mixture probability component $\lambda_{isi,g_1,g_2,g_3}$. This is done via introducing latent cluster allocation variables $\{z_{isi,r,w}\}_{r=1,w=1}^{3,d_r}$, as before, with $z_{isi,r,w}$ indicating the cluster label for the wth level of the rth covariate. Importantly, as in the case of transition probabilities, when the levels of a covariate r are all clustered together, that is, $k_{isi,r} = 1$, the covariate r has no effect on the ISI distribution, allowing us to easily and formally test its significance based on the posterior probability of the event $k_{isi,r} = 1$.

The ISI mixture probability in a song s produced by a mouse $i_s = i$ with genotype $x_{s,1}$ in cluster g_1 and context $x_{s,2}$ in cluster g_2 and syllable pair $(y_{s,t-1}, y_{s,t})$ in cluster g_3 is given by

$$\mathbf{P}_{isi,g_1,g_2,g_3}^{(i)}(\cdot) = \pi_{isi,0}^{(i)}(\cdot)\lambda_{isi,g_1,g_2,g_3}(\cdot) + \pi_{isi,1}^{(i)}(\cdot)\lambda_{isi}^{(i)}(\cdot).$$

As earlier, we assign conditionally conjugate Dirichlet priors to the fixed and random effect components and give a Beta prior to the mouse-specific coefficient,

$$\lambda_{isi,g_1,g_2,g_3}(\cdot) \sim \text{Dir}\left\{\alpha_{isi,0}\lambda_{isi,0}(1), \ldots, \alpha_{isi,0}\lambda_{isi,0}(K)\right\},$$
$$\lambda_{isi}^{(i)}(\cdot) \sim \text{Dir}\left\{\alpha_{isi}^{(0)}\lambda_{isi,0}(1), \ldots, \alpha_{isi}^{(0)}\lambda_{isi,0}(K)\right\}, \quad \pi_{isi,0}^{(i)}(k) \sim \text{Beta}(a_{isi,0}, a_{isi,1}).$$

The random effects $\lambda_{isi}^{(i)}$ and $\pi_{isi,0}^{(i)}$ can be easily integrated out to obtain the closed-form population-level mixture probabilities as

$$P_{isi,g_1,g_2,g_3}(k) = \pi_{isi,0}\lambda_{isi,g_1,g_2,g_3}(k) + \pi_{isi,1}\lambda_{isi,0}(k),$$

where $\pi_{isi,0} = \frac{a_{isi,0}}{a_{isi,0}+a_{isi,1}}$ and $\pi_{isi,1} = \frac{a_{isi,1}}{a_{isi,0}+a_{isi,1}}$. Finally, we let $\lambda_{isi,0}$ center around a global $\lambda_{isi,00}$ and the hyperparameters $\alpha_{isi}^{(0)}$ and $\alpha_{isi,0}$ are given gamma hyperpriors.

The complete Bayesian hierarchical model for the ISIs may be summarized as

$$(\widetilde{\tau}_{s,t} \mid z_{isi,s,t} = k) \sim \text{Ga}(\widetilde{\tau}_{s,t} \mid \alpha_k, \beta_k),$$
$$(z_{isi,s,t} \mid i_s = i, z_{isi,1,x_{s,1}} = g_1, z_{isi,2,x_{s,2}} = g_2, z_{isi,3,(y_{s,t-1},y_{s,t})} = g_3) \sim$$
$$\text{Mult}\left\{P_{isi,g_1,g_2,g_3}^{(i)}(1), \ldots, P_{isi,g_1,g_2,g_3}^{(i)}(K)\right\},$$
$$P_{isi,g_1,g_2,g_3}^{(i)}(k) = \pi_{isi,0}^{(i)}(k)\lambda_{isi,g_1,g_2,g_3}(k) + \pi_{isi,1}^{(i)}(k)\lambda_{isi}^{(i)}(k), \tag{3.5}$$
$$\lambda_{isi}^{(i)}(\cdot) \sim \text{Dir}\left\{\alpha_{isi}^{(0)}\lambda_{isi,0}(1), \ldots, \alpha_{isi}^{(0)}\lambda_{isi,0}(K)\right\}, \quad \alpha_{isi}^{(0)} \sim \text{Ga}(a_{isi}^{(0)}, b_{isi}^{(0)}),$$
$$\lambda_{isi,g_1,g_2,g_3}(\cdot) \sim \text{Dir}\left\{\alpha_{isi,0}\lambda_{isi,0}(1), \ldots, \alpha_{isi,0}\lambda_{isi,0}(K)\right\}, \quad \alpha_{isi,0} \sim \text{Ga}(a_{isi,0}, b_{isi,0}),$$
$$\lambda_{isi,0}(\cdot) \sim \text{Dir}\left\{\alpha_{isi,00}\lambda_{isi,00}(1), \ldots, \alpha_{isi,00}\lambda_{isi,00}(K)\right\},$$
$$\pi_{isi,0}^{(i)}(k) \sim \text{Beta}(a_{isi,0}, a_{isi,1}), \quad \alpha_k \sim \text{Ga}(a_{isi,0}, b_{isi,0}), \quad \beta_k \sim \text{Ga}(a_{isi,0}, b_{isi,0}.)$$

Gamma mixtures, in other forms, have appeared before in Chen (2000), Wiper *and others* (2001), Hanson (2006), etc. Such mixtures can approximate a large class of distributions on $[0, \infty)$ (see, e.g., Theorem 14 in Wu and Ghosal, 2008) and hence allow a flexible nonparametric estimation of the ISI distributions. The gamma kernel, however, brings in computational challenges which we address briefly in Section 4 below and then in detail again in Appendix D of the Supplementary material available at *Biostatistics* online. There exists some literature on flexible mixture and partition models for conditionally varying densities of continuous random variables in the presence of covariates (MacEachern, 1999; Chung and Dunson, 2009; Müller *and others*, 2011, etc.). To our knowledge, however, flexible mixed effects mixture models of the type proposed here that accommodate both fixed effects of covariates as well as random heterogeneity of subjects while also allowing simultaneous covariate selection have not appeared in the literature before.

4. Prior hyperparameters and posterior inference

The choices of prior hyperparameters, including the choice of the number of mixture components K in the gamma mixture model (3.3) for which we use predictive model selection criteria, are discussed in Appendix B of the Supplementary material available at *Biostatistics* online.

Our posterior inference is based on samples of the model parameters drawn using an MCMC algorithm. The full conditional posterior distributions are mostly obtained in closed-form and are easy to sample from. One exception is the sampling of the gamma mixture parameters. The conjugate prior for gamma distribution is known to be analytically intractable (Damsleth, 1975; Miller, 1980), posing difficulty in sampling the parameters α_k's and β_k's. We experimented with a number of ideas and ultimately used the strategy introduced in Miller (2019) which uses a gamma density function to approximate the full conditional for gamma shape parameters. After sampling the shape parameters α_k's, the rate parameters β_k's can be easily sampled from their closed-form conjugate gamma full conditionals. As the results of Sections 5 and Appendix F of the Supplementary material available at *Biostatistics* online will illustrate, this method worked well with real data as well as in our simulation experiments, converging quickly, mixing well and providing accurate estimates of the target distributions. Details of the posterior sampling algorithm can be found in Appendix D of the Supplementary material available at *Biostatistics* online.

5. Results for the Foxp2 data set

In this section, we discuss the results of the proposed Bayesian Markov renewal mixed model fitted to the Foxp2 data set. We present the results for syllable transitions and ISIs separately.

5.1. *Results for syllable transition*

Figure 3a shows the estimated posterior mean of the population-level transition probabilities, $P_{\text{trans},x_1,x_2}(y_t \mid y_{t-1})$, given genotype x_1 and social context x_2. We see that regardless of the covariate values, the s syllable is predominantly transitioned to and u is the least likely syllable to transition to across different genotypes and contexts. This is reasonable since the s syllable is presumably the easiest to pronounce and u is the least pronounced syllable across all sequences.

We evaluate the global impact of covariates by computing the estimated posterior distribution of $k_{\text{trans},j}$ for genotype ($j = 1$) and social context ($j = 2$) (Figure S3 in Appendix E of the Supplementary material available at *Biostatistics* online). The estimated posterior probability that $k_{\text{trans},1}$ greater than 1 is approximately 0.56. In contrast to the findings reported previously in Sarkar *and others* (2018), the evidence that the mutation on the Foxp2 gene impacts the transition probabilities of the syllables
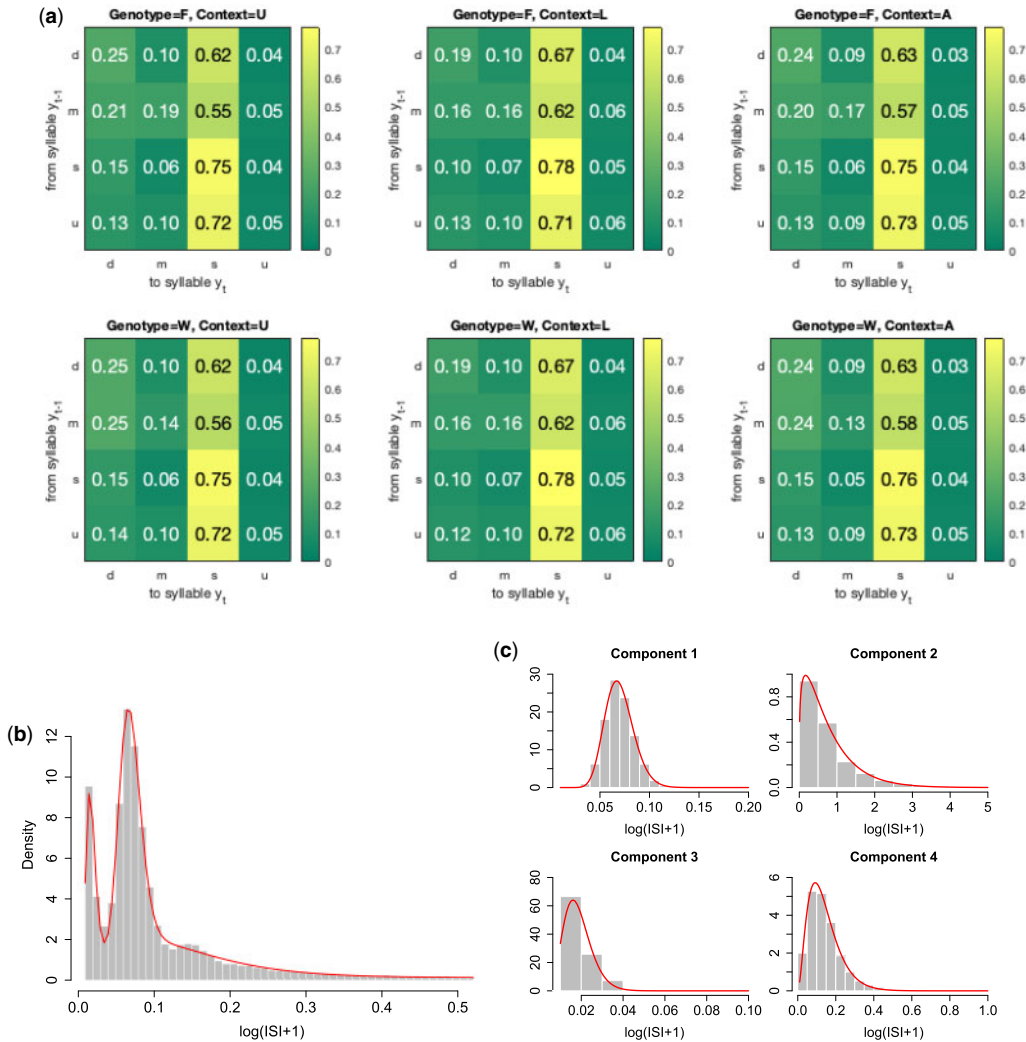
Fig. 3. Results for the Foxp2 data set. (a) The estimated posterior mean of the transition probabilities $P_{trans,x_1,x_2}(y_t \mid y_{t-1})$ of syllables $y_t, y_{t-1} \in \mathcal{Y} = \{d, m, s, u\}$ under different combinations of genotype $x_1 \in \{F, W\}$ and social context $x_2 \in \{U, L, A\}$. (b) Histogram of the transformed ISIs with the estimated posterior mean (red line) of their marginal gamma mixture density based on MCMC samples after burn-in and thinning. (c) Histograms of the transformed ISIs for each component of the gamma mixture model along with the component density (red lines) from the last MCMC iteration. The x-axes are adjusted for better visualization.

has thus become somewhat weaker. This illustrates that treating ISI as a separate continuous variable results in less distinctions between the transition dynamics of the two groups. Conforming to the previous analyses in Sarkar *and others* (2018), there is, however, very clear evidence of the influence of social context on the transitions probabilities. Specifically, whenever there were two clusters, the contexts U and A were clustered together. In Figure 3a, we see that there is strong evidence that contexts U and A have similar impact across genotypes. Compared to U and A, the L context has smaller transition probabilities for transitions types $d \rightarrow d$ and $m \rightarrow d$ across the two genotypes.

The decrease in the transition probabilities of d → d and m → d seems to be explained by the increase in the probabilities of d → s and m → s, suggesting that the mice vocalized short and simple symbols more often under the *L* context.

In Sarkar *and others* (2018), the coefficients $(\pi_{\text{trans},0}, \pi_{\text{trans},1})$ were assumed to be shared between all mice, whereas here we have allowed them to be mouse specific as $(\pi_{\text{trans},0}^{(i)}, \pi_{\text{trans},1}^{(i)})$. To investigate the effectiveness of this, we looked at the values for these coefficients for different mice. Given a preceding syllable y_{t-1}, we compare the coefficient $\pi_{\text{trans},0}^{(i)}$ for each individual i. We present these results in Table S3 in Appendix E of the Supplementary material available at *Biostatistics* online. From the table, we see that these coefficients differ substantially between different mice, especially for preceding syllables u and m, justifying our decision of making them mouse specific.

5.2. *Results for ISIs*

We first transform the observed ISIs to $\widetilde{\tau}_{s,t} = \log(1 + \tau_{s,t})$. The original $\tau_{s,t}$'s have a wide range with a minimum of 0.01 and a maximum of 258.8 s. This preprocessing step helps shorten the range of the data and produce better graphical summaries for the results. We add 1 to the original $\tau_{s,t}$'s before taking the log to avoid negative $\widetilde{\tau}_{s,t}$'s.

Figure 3b shows the histogram of the transformed ISIs along with the posterior mean (red curve) averaged from samples after burn-ins and thinning. Table 1c in Section 2 shows the empirical and posterior means and standard deviations of the ISIs side-by-side. We see that our proposed model fits the ISI data very well, whose two peaks are captured by the mixture gamma distribution. Figure 3c displays the histogram for each mixture component along with the corresponding gamma densities with shape and rate parameters taken from the last iteration of the MCMC sampler. It is clear that components 1 and 3 represent the two peaks we see in Figure 3b whereas components 2 and 4 assign more probability mass to larger values of the transformed ISIs. The parameters of each component gamma distribution are presented in the first table in Table 2. The shape and rate parameters for components 1 and 3 are much larger than those of components 2 and 4, capturing the concentration of the small values of the transformed ISIs.

Table 2 displays the mixture probabilities taken from the last MCMC iteration for each covariate: genotype (*F* and *W*), social contexts (*U*, *L*, and *A*), and every preceding–current syllable pair. We see that mice with the Foxp2 mutation have a smaller mixture probability in the components 1, 3, and 4 compared to wild-types but have a significantly higher probability for component 2 (+0.08). Recall that component 2 has the smallest rate parameter, which indicates a higher probability to have a larger value compared to the other components. A large mixture probability in component 2 indicates a high ISI value. This suggests that the ISI length for mice with the Foxp2 mutation concentrates on larger values than the ISIs of wild-types, which supports our hypothesis that mice with such mutation needs a longer ISI before pronouncing a new syllable. One interesting discovery from Tables 1c and 2 is that male mice in the presence of a live female (*L* context) tend to have shorter ISIs than those under the other two contexts since the mixture probability is much higher in component 3 for the *L* context compared to the other contexts. This suggests that male mice need a shorter interval between pronouncing two syllables in the presence of a live female. This finding, along with the discovery that transitions to the simplest syllable s are more frequent under the *L* context, shows that male mice exhibit different vocalization dynamics when there is an awake female mouse present.

The last two tables in Table 2 show the mixture probabilities associated with each preceding-current syllable pair. The 16 pairs of syllables have similar weights in the four components except for (s, s), which has a much larger mixture probability for component 2, indicating a longer ISI between pronouncing consecutive s syllables. This corresponds to the result in Table 1c where the ISIs are

Table 2. *Results for ISIs for the Foxp2 data set taken from the last MCMC iteration*

Estimated gamma shape and rate parameters			Estimated mixture probabilities for each genotype			Estimated mixture probabilities for each context		
	shape.k	rate.k		F	W		U	L	A
Comp 1	23.47	336.41	Comp 1	0.52	0.57	Comp 1	0.56	0.42	0.65
Comp 2	1.28	1.69	Comp 2	0.16	0.08	Comp 2	0.15	0.11	0.09
Comp 3	7.79	423.98	Comp 3	0.08	0.11	Comp 3	0.05	0.18	0.05
Comp 4	2.89	20.59	Comp 4	0.24	0.25	Comp 4	0.23	0.29	0.21

Estimated mixture probabilities for each preceding–current syllable pair

	(d, d)	(d, m)	(d, s)	(d, u)	(m, d)	(m, m)	(m, s)	(m, u)
Comp 1	0.65	0.65	0.45	0.47	0.65	0.65	0.47	0.55
Comp 2	0.07	0.07	0.16	0.12	0.07	0.07	0.12	0.14
Comp 3	0.05	0.05	0.10	0.17	0.05	0.05	0.17	0.07
Comp 4	0.23	0.23	0.29	0.24	0.23	0.23	0.24	0.24

Estimated mixture probabilities for each preceding–current syllable pair (cont'd)

	(s, d)	(s, m)	(s, s)	(s, u)	(u, d)	(u, m)	(u, s)	(u, u)
Comp 1	0.56	0.65	0.35	0.55	0.47	0.54	0.47	0.56
Comp 2	0.14	0.07	0.23	0.14	0.12	0.11	0.12	0.14
Comp 3	0.06	0.05	0.11	0.07	0.17	0.11	0.17	0.06
Comp 4	0.24	0.23	0.31	0.24	0.24	0.24	0.24	0.24

usually longer when both the preceding and the current syllable are s, no matter what the genotype or social context is. Additional results on ISIs, including results of global tests for genotype and context (Figure S4 in Appendix E of the Supplementary material available at *Biostatistics* online), are provided in the supplementary materials.

6. DISCUSSION

This article introduced a new class of Bayesian Markov renewal mixed effects models that allows inference of both state transition probabilities and continuous interstate interval times. On the statistical side, our main novel contribution is a mixed effects gamma mixture model for the inter-state intervals. The mixture probabilities build on carefully constructed convex combinations of a fixed effect component for the associated covariates and a random effect component for the associated individual, resulting in a highly flexible and computationally tractable model. At the same time, covariate values that induce similar effects on the response are probabilistically clustered together and, in the process, significant covariates are identified.

We used the model to reanalyze the Foxp2 data set which comprises a collection of songs sung by adult male mice with or without a Foxp2 mutation under various social contexts. In contrast to previous analyses, we found weaker evidence that the transition dynamics of the syllables within the songs vary with genotype and social context. On the other hand, there is significant evidence that all three covariates, namely genotypes, social contexts, and preceding–current syllable pairs, influence the lengths of the ISIs. The important scientific implication is that the vocal impairment

of the Foxp2 mice is manifested in their having longer ISIs than wild-types, not just in the syllable transition dynamics as previous analyses suggested.

The mixture gamma model for the ISIs that we proposed here may be of independent interest outside the scope of the Foxp2 application. The model is nontrivial and brought in additional statistical challenges, including posterior computation for unknown gamma parameters and selection of the unknown number of mixture components. To our knowledge, sophisticated mixed effects mixture models for continuous variables that also simultaneously allow covariate selection have not been proposed in the literature before.

Our model is quite generic in nature and hence can be used to analyze other data sets comprising categorical sequences and associated continuous interstate interval times, both of which may potentially be influenced by various exogenous factors as well as subject-specific heterogeneity. The field of vocal communication neuroscience constitutes an important area of neuroscience research. Investigating scientific questions using animal models in controlled laboratory environments is a standard practice in this field. These researchers often use the standard two-way analysis of variance (ANOVA) design like the Foxp2 study analyzed in our manuscript. Our method is broadly applicable to such studies. We also cite below some examples from other application domains that can benefit from such analyses. In a study of asthma patients, Combescure *and others* (2003) estimated the control states (optimal, suboptimal or unacceptable) of 371 asthma patients with different BMI and disease severity over a 4-year period. In an education assessment study, Zhang *and others* (2019) recorded sequences of writing states, characterized by keystroke logs, for 257 eighth graders of various genders, races, and socioeconomic statuses. These works used Cox regression models to incorporate potential covariate influences but ignored the heterogeneity between the individuals (Dabrowska *and others*, 1994; Król and Saint-Pierre, 2015; Guo *and others*, 2019). To our knowledge, there exists no other statistical approach that flexibly accommodates population and individual-level effects in the transition as well as the interstate interval distributions while also selecting the significant covariates for both.

Finally, in this article, we focused on discrete exogenous factors due to the nature of our motivating Foxp2 application. A simple but practically effective way to incorporate continuous covariates into the model is by categorizing them using, for example, their quantiles. Future research could investigate more principled ways to incorporate continuous covariates in our model. The proposed model also points to possible directions of research in Markov renewal models that consider different aspects of state transitions (transition probabilities, interstate intervals, state durations, etc.) and help make novel discoveries in important practical applications.

Software

An R package BMRMM implementing our method is available at the Comprehensive R Archive Network (CRAN) and can also be accessed at https://github.com/abhrastat/BMRMM. A "readme" file is included here that instructs how to use the package to run the specific analysis reported here.

Supplementary material

Supplementary material is available at http://biostatistics.oxfordjournals.org.

Acknowledgments

Conflict of Interest: None declared.

REFERENCES

ALVAREZ, E. E. (2005). Estimation in stationary Markov renewal processes, with application to earthquake forecasting in Turkey. *Methodology and Computing in Applied Probability* **7**, 119–130.

ARRIAGA, G. AND JARVIS, E. D. (2013). Mouse vocal communication system: are ultrasounds learned or innate? *Brain and Language* **124**, 96–116.

BULLA, P. AND MULIERE, P. (2007). Bayesian nonparametric estimation for reinforced Markov renewal processes. *Statistical Inference for Stochastic Processes* **10**, 283–303.

CASTELLUCCI, G. A., MCGINLEY, M. J. AND MCCORMICK, D. A. (2016). Knockout of Foxp2 disrupts vocal development in mice. *Nature Scientific Reports* **6**, 1–14. doi: 10.1038/srep23305.

CHABOUT, J., SARKAR, A., DUNSON, D. B. AND JARVIS, E. D. (2015). Male song syntax depends on contexts and influences female preferences in mice. *Frontiers in Behavioral Neuroscience* **9**, 1–19.

CHABOUT, J., SARKAR, A., PATEL, S., RAIDEN, T., DUNSON, D. B., FISHER, S. E. AND JARVIS, E. D. (2016). A Foxp2 mutation implicated in human speech deficits alters sequencing of ultrasonic vocalizations in adult male mice. *Frontiers in Behavioral Neuroscience* **10**, 1–18.

CHABOUT, J., SERREAU, P., EY, E., BELLIER, L., AUBIN, T., BOURGERON, T. AND GRANON, S. (2012). Adult male mice emit context-specific ultrasonic vocalizations that are modulated by prior isolation or group rearing environment. *PLoS One* **7**, e29401.

CHEN, S. X. (2000). Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics* **52**, 471–480.

CHUNG, Y. AND DUNSON, D. B. (2009). Nonparametric Bayes conditional distribution modeling with variable selection. *Journal of the American Statistical Association* **104**, 1646–1660.

COMBESCURE, C., CHANEZ, P., SAINT-PIERRE, P., DAURES, J. P., PROUDHON, H., GODARD, P. *and others*. (2003). Assessment of variations in control of asthma over time. *European Respiratory Journal* **22**, 298–304.

DABROWSKA, D. M., SUN, G.-W. AND HOROWITZ, M. M. (1994). Cox regression in a Markov renewal model: an application to the analysis of bone marrow transplant data. *Journal of the American Statistical Association* **89**, 867–877.

DAMSLETH, E. (1975). Conjugate classes for Gamma distributions. *Scandinavian Journal of Statistics* **2**, 80–84.

EPIFANI, I., LADELLI, L. AND PIEVATOLO, A. (2014). Bayesian estimation for a parametric Markov renewal model applied to seismic data. *Electronic Journal of Statistics* **8**, 2264–2295.

FISHER, S. E. AND SCHARFF, C. (2009). Foxp2 as a molecular window into speech and language. *Trends in Genetics* **25**, 166–177.

FOUCHER, Y., MATHIEU, E., SAINT-PIERRE, P., DURAND, J.-F. AND DAURÈS, J.-P. (2005). A semi-Markov model based on generalized Weibull distribution with an illustration for HIV disease. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* **47**, 825–833.

FUJITA, E., TANABE, Y., SHIOTA, A., UEDA, M., SUWA, K., MOMOI, M. Y. AND MOMOI, T. (2008). Ultrasonic vocalization impairment of Foxp2 (R552H) knockin mice related to speech-language disorder and abnormality of Purkinje cells. *Proceedings of the National Academy of Sciences United States of America* **105**, 3117–3122.

GAUB, S., FISHER, S. E. AND EHRET, G. (2016). Ultrasonic vocalizations of adult male Foxp2 mutant mice: behavioral contexts of arousal and emotion. *Genes, Brain and Behavior* **15**, 243–259.

GUO, H., ZHANG, M., DEANE, P. AND BENNETT, R. E. (2019). Writing process differences in subgroups reflected in keystroke logs. *Journal of Educational and Behavioral Statistics* **44**, 571–596.

HANSON, T. E. (2006). Modeling censored lifetime data using a mixture of gammas baseline. *Bayesian Analysis* **1**, 575–594.

HOLY, T. E. AND GUO, Z. (2005). Ultrasonic songs of male mice. *PLoS Biology* **3**, 2177–2186.

JARVIS, ERICH D. (2019). Evolution of vocal learning and spoken language. *Science* **366**, 50–54.

KRÓL, A. AND SAINT-PIERRE, P. (2015). SemiMarkov: an R package for parametric estimation in multi-state semi-Markov models. *Journal of Statistical Software* **66**, 1–16.

LAI, C. S. L., FISHER, S. E., HURST, J. A., VARGHA-KHADEM, F. AND MONACO, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519–523.

LEVY, P. (1954). Processus semi-Markoviens. In: *Proceedings of International Congress of Mathematics*.

MACEACHERN, S. N. (1999). Dependent nonparametric processes. In: *ASA Proceedings of the Section on Bayesian Statistical Science*. Alexandria, Virginia: American Statistical Association, pp. 50–55.

MILLER, J. W. (2019). Fast and accurate approximation of the full conditional for Gamma shape parameters. *Journal of Computational and Graphical Statistics* **28**, 476–480.

MILLER, R. B. (1980). Bayesian analysis of the two-parameter Gamma distribution. *Technometrics* **22**, 65–69.

MOLES, A., COSTANTINI, F., GARBUGINO, L., ZANETTINI, C. AND DÁMATO, F. R. (2007). Ultrasonic vocalizations emitted during dyadic interactions in female mice: a possible index of sociability? *Behavioural Brain Research* **182**, 223–230.

MOONEY, R. (2020). The neurobiology of innate and learned vocalizations in rodents and songbirds. *Current Opinion in Neurobiology* **64**, 24–31.

MULIERE, P., SECCHI, P. AND WALKER, S. G. (2003). Reinforced random processes in continuous time. *Stochastic Processes and Their Applications* **104**, 117–130.

MÜLLER, P., QUINTANA, F. AND ROSNER, G. L. (2011). A product partition model with regression on covariates. *Journal of Computational and Graphical Statistics* **20**, 260–278.

MUSOLF, K., MEINDL, S., LARSEN, A. L., KALCOUNIS-RUEPPELL, M. C. AND J., PENN D. (2015). Ultrasonic vocalizations of male mice differ among species and females show assortative preferences for male calls. *PLoS One* **10**, e0134123.

NIH-NIDCD REPORT. (2020). Statistics on voice, speech, and language. https://www.nidcd.nih.gov/health/statistics/statistics-voice-speech-and-language.

PHELAN, M. J. (1990). Bayes estimation from a Markov renewal process. *The Annals of Statistics* **18**, 603–616.

PYKE, R. (1961). Markov renewal processes: definitions and preliminary properties. *The Annals of Mathematical Statistics* **32**, 1231–1242.

SARKAR, A., CHABOUT, J., MACOPSON, J. J., JARVIS, E. D. AND DUNSON, D. B. (2018). Bayesian semiparametric mixed effects Markov models with application to vocalization syntax. *Journal of the American Statistical Association* **113**, 1515–1527.

SCATTONI, M. L., RICCERI, L. AND CRAWLEY, J. N. (2011). Unusual repertoire of vocalizations in adult BTBR T+tf/J mice during three types of social encounters. *Genes, Brain and Behavior* **10**, 44–56.

SMITH, W. L. (1955). Regenerative stochastic processes. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **232**, 6–31.

VARGHA-KHADEM, F., WATKINS, K. E., PRICE, C. J., ASHBURNER, J., ALCOCK, K. J., CONNELLY, A., FRACKOWIAK, R. S. J., FRISTON, K. J., PEMBREY, M. E., MISHKIN, M. *and others*. (1998). Neural basis of an inherited speech and language disorder. *Proceedings of the National Academy of Sciences United States of America* **95**, 12695–12700.

Weiss, G. H. and Zelen, M. (1965). A semi-Markov model for clinical trials. *Journal of Applied Probability* **2**, 269–285.

Wiper, M., Insua, D. R. and Ruggeri, F. (2001). Mixtures of gamma distributions with applications. *Journal of Computational and Graphical Statistics* **10**, 440–454.

Wu, Y. and Ghosal, S. (2008). Kullback–Leibler property of kernel mixture priors in Bayesian density estimation. *Electronic Journal of Statistics* **2**, 298–331.

Yang, M. C. K. and Hursch, C. J. (1973). The use of a semi-Markov model for describing sleep patterns. *Biometrics* **29**, 667–676.

Zhang, M., van Rijn, P. W., Deane, P. and Bennett, R. E. (2019). Scenario-based assessments in writing: an experimental study. *Educational Assessment* **24**, 73–90.