Presentation of Ancestral Reconstruction Results

J.P. Meagher

8 January 2018

Introduction

We are given a dataset of N bat echolocation call recordings denoted $\{y_n\}_{n=1}^{N}$. This recording is then processed to produce a set of smooth surfaces over a regular grid denoted $\{\tilde{S}_n\}_{n=1}^{N}$. This surface is produced by smoothing the call spectrogram and mapping it to a regular grid over relevant frequencies and an absolute time scale Pigoli et al. (2015).

Along with this dataset we are given a phylogeny defining the evolutionary relationships between the species of bat.

A model has been developed to produce ancestral reconstructions for the smoothed spectrogram surfaces representing the echolocation calls of extinct bats and an audio file approximating the call that would correspond to such a spectrogram surface can also be produced.

Thus, for the dataset of Mexican Bat echolocation calls and the given Phylogeny, Ancestral Reconstruction has been performed.

The Current Model

An illustration of the current iteration of a model for the evolution of bat echolocation calls is presented in Figure 4.

A Joint model for spectrogram surfaces

Pigoli, Davide, Pantelis Z Hadjipantelis, John S Coleman, and John AD Aston. 2015. "The Analysis of Acoustic Phonetic Data: Exploring Differences in the Spoken Romance Languages." *arXiv Preprint arXiv:1507.07587.*
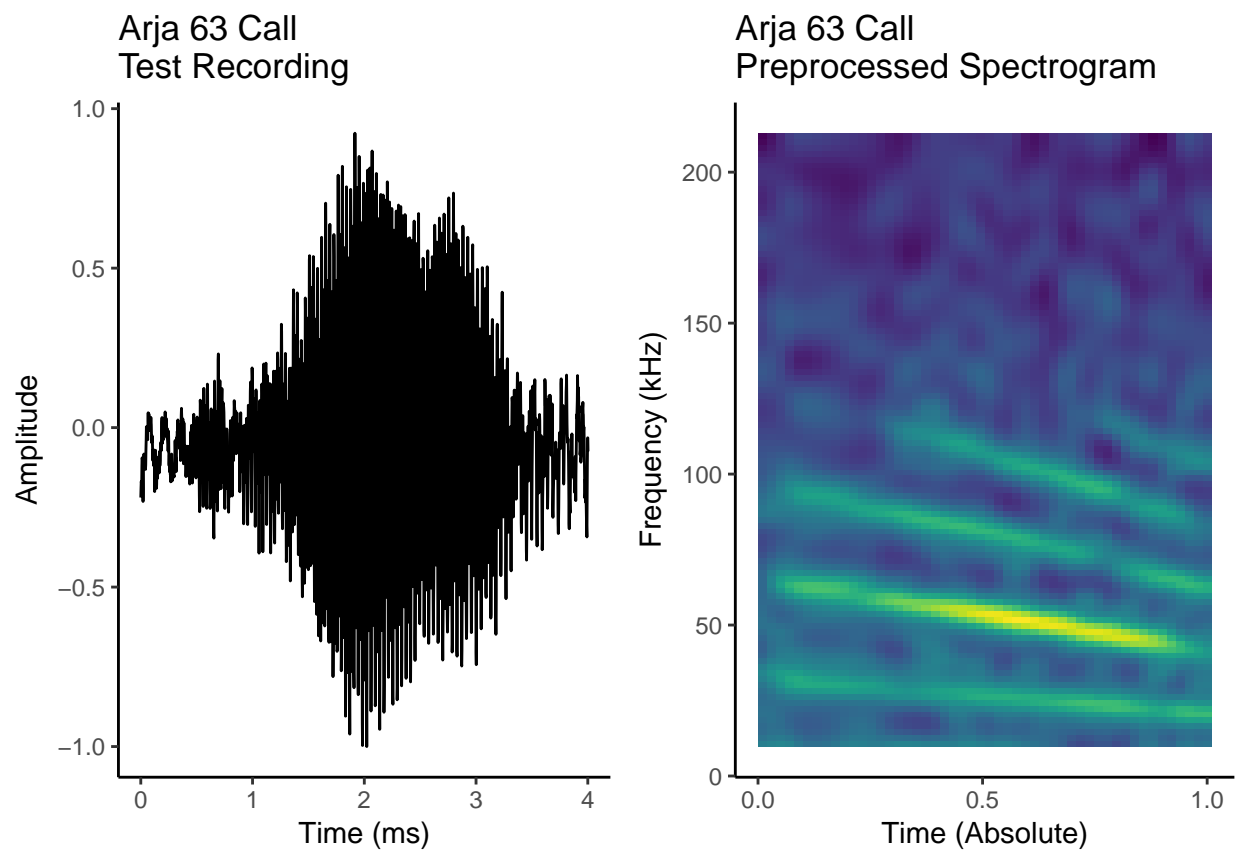
Figure 1: A randomly selected bat call from the species Arja alongside it's corresponding smoothed surface representation. The smooth surface is obtained by taking the call spectrogram and treating it as a functional data object. The spectrogram is first smoothed by a robust 2-D spline smoother, then mapped to an absolute time scale and registered in time by a pairwise surface synchronisation, and finally restricted to the 9 - 212 kHz frequency spectrum.
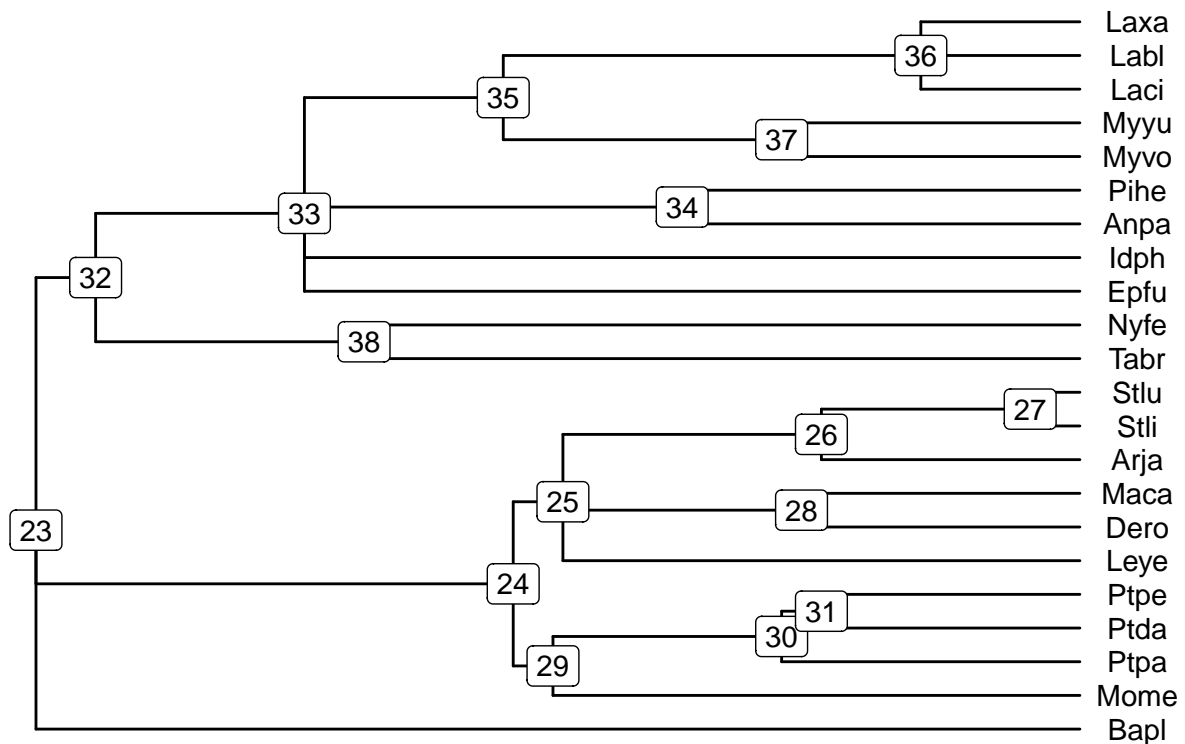
Bat Phylogenetic Tree

Figure 2: Tree of assumed evolutionary relationships between Bat Species. This phylogeny was transcribed from a recent bat super tree and shoud represent a 'best guess' for the evolutionary relationships between bat species based on the fossil record alongside morphological and molecular studies of evolutionary relationships.
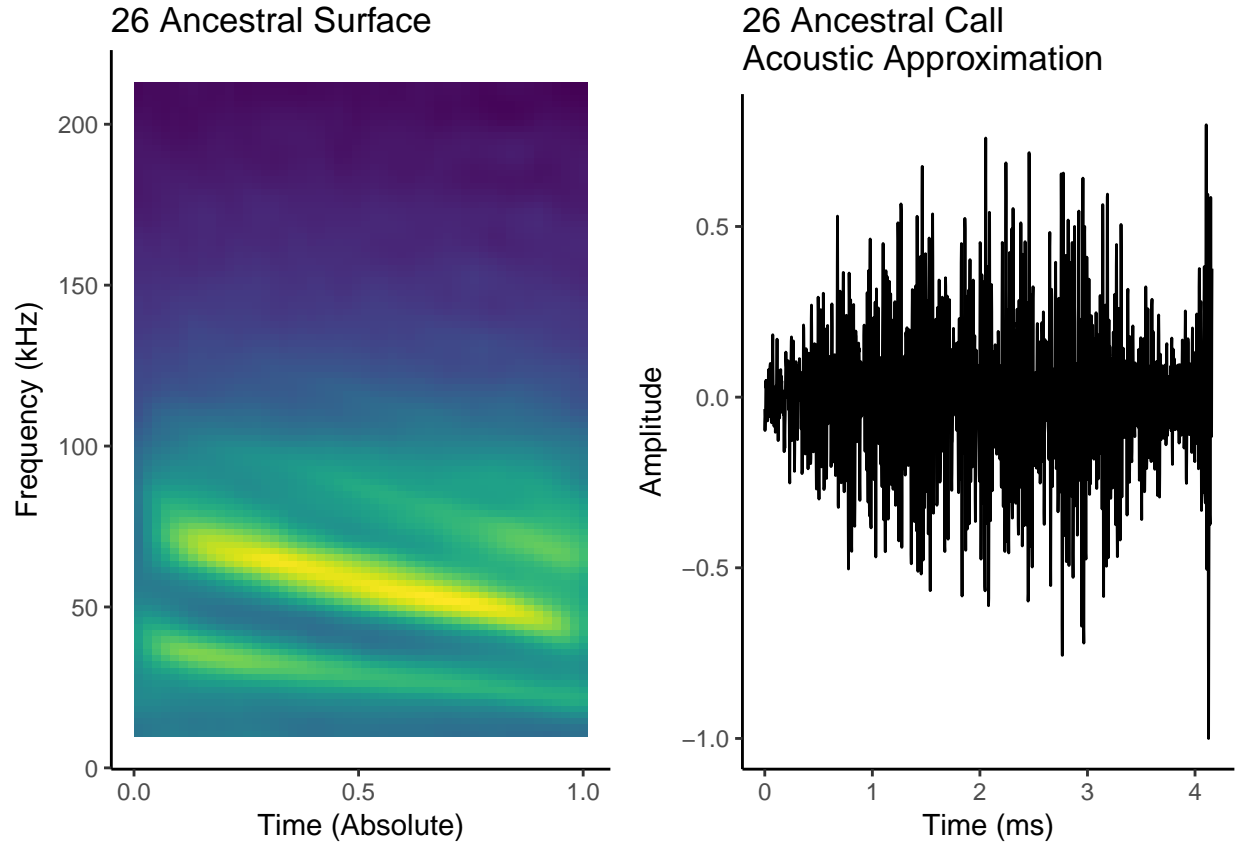
Figure 3: Ancestral spectrogram and acoustic approximation for the common ancestor of Arja, Stli, Stlu. The reconstruction is given by the MAP estimates for the weight of each evolutionary feature at the node. In this case, evolutionary features were identified by a PCA of the smoothed spectrogram surfaces. The acoustic reconstruction was performed assuming a call duration of approximately 4 ms
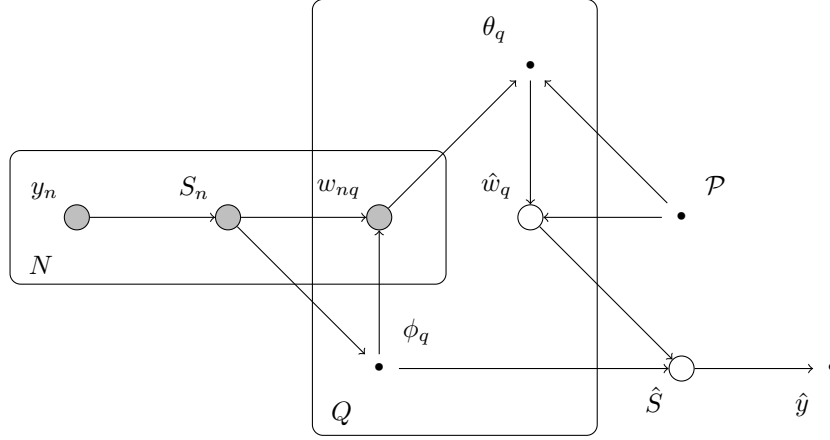
Figure 4: A Graphical model detailing the structure of the model for evolution used to produce reconstructions of ancestral bat echolocation calls. Let y_n be a random variable representing an echolocation call recording. S_n is the random variable representing the smoothed spectrogram surface given by y_n. The Mexican bat call dataset provides $N = 1816$ observations of these random variables. The process of transforming a call recording into a spectrogram surface was covered in my 9 month report. The model assumes that each S_n can be modelled by Q independent deterministic 'evolutionary features' denoted ϕ_q. In this case ϕ_q is inferred by a Principal Components Analysis of $\{S_n\}_{n=1}^N$. The weight of each evolutionary feature in S_n is itself a random variable, where w_{nq} denotes the weight of ϕ_q in S_n. w_{nq} is assumed to behave as an Ornstein-Uhlenbeck Gaussian process for which the input space is the phylogeny $\mathcal{P}$. Each Gaussian process is defined by the deterministic hyperparameters $\theta_q = [\gamma_q, \ell_q, \sigma_q]^\mathsf{T}$ which are inferred from the data by Type II maximum likelihood estimation over the observed weights. The phylogeny $\mathcal{P}$ is also assumed to be deterministic in this model and is shown in Figure 2. Ancestral reconstruction is performed by making a prediction for the feature weights, denoted $\hat{w}$, at some point on $\mathcal{P}$. Applying these weights to the evolutionary features produces the ancestral call surface which in turn provides an estimate for the ancestral call.
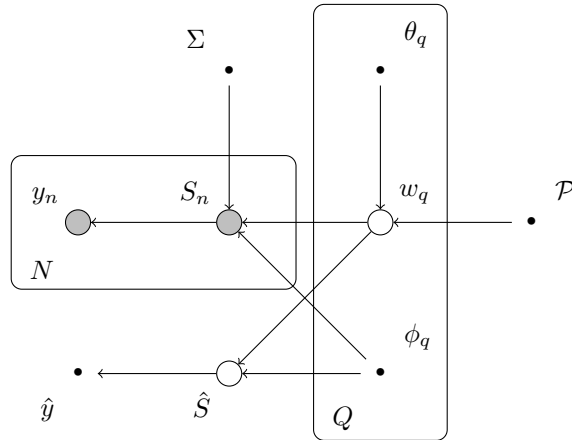

Figure 5: A proposed extension to the model for the evolution a bat echolocation calls presented in Figure 4. This model includes a noise process over the call surface, which would allow the calculation of the model evidence. This in turn would facilitate model selection for various sets of evolutionary features, phylogenies, and Ornstein-Uhlenbeck process hyperparameters.