Current Biology, Volume *31*

Supplemental Information

Neurally driven synthesis

of learned, complex vocalizations

Ezequiel M. Arneodo, Shukai Chen, Daril E. Brown II, Vikash Gilja, and Timothy Q. Gentner

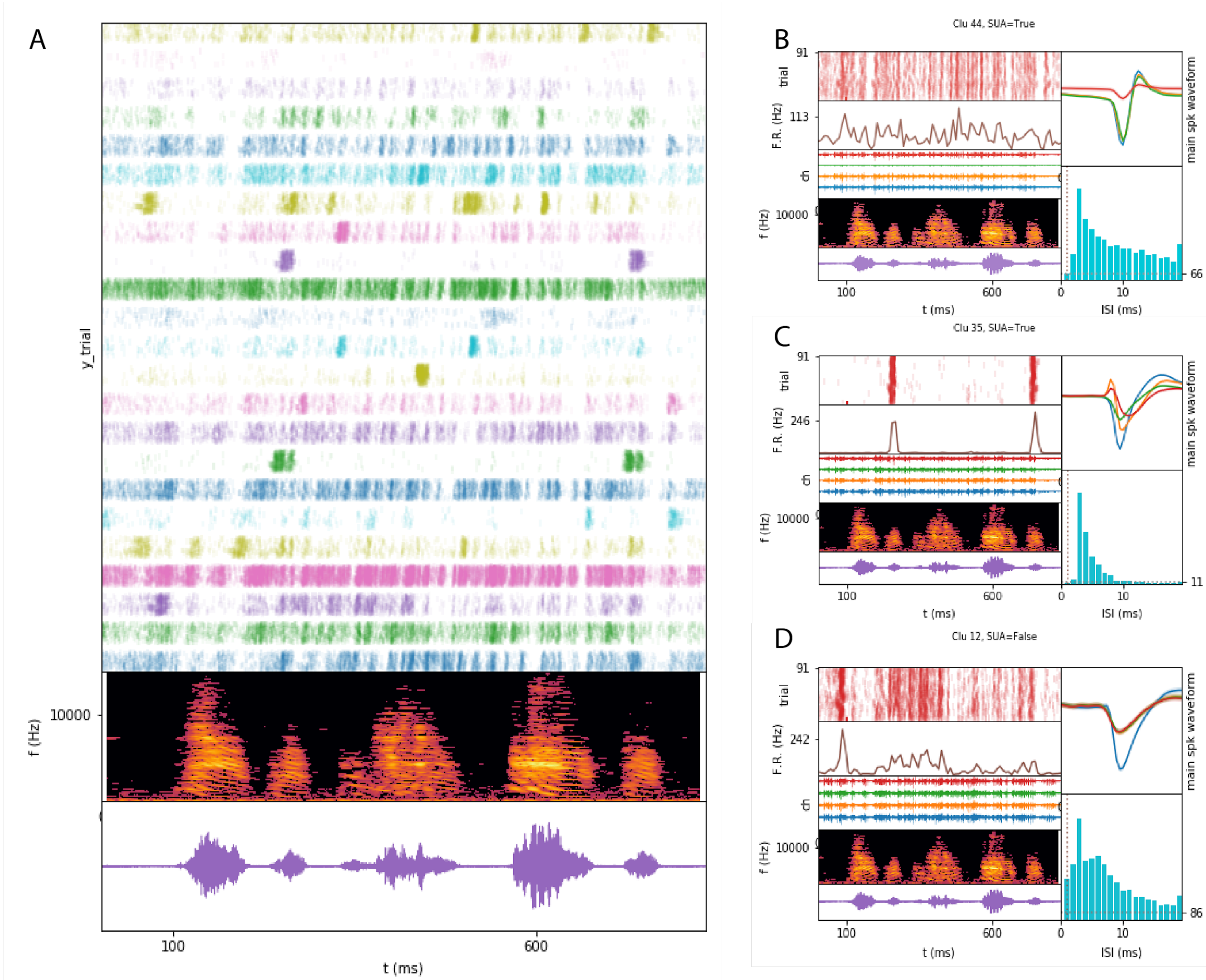

Figure S1. Example clusters of sorted units, Related to Figure 1

(A) Raster plots of 23 automatically clustered units, for 91 repetitions of a motif, spanning 12 hours of recording (top), aligned to an example motif's spectrogram (middle) and waveform (bottom). **(B)** Example of a (putatively) single unit activity cluster (SUA), likely an interneuron (HVC_I). In the left panels are, from bottom to top: the raster; the corresponding histogram (10ms bin); example traces of the 4 neural channels where the cluster's representative waveform has the largest amplitudes; spectrogram of an example motif; waveform of an example motif. In the right panels: (top) a representative waveform for the cluster (mean of 10,000 events), plotted for the 4 channels with the largest amplitude (peak to trough) and (bottom) inter-spike-interval (ISI) histogram (0.5ms bins). Vertical dotted line indicates 1ms, horizontal dotted line indicates 3% level of refractory period violations. **(C)** Example of a (putatively) SUA cluster, likely a projection neuron (HVC_x or HVC_{RA}). **(D)** Example of a (putatively) multi unit activity cluster (MUA). For this study, we used the highest yield session for each bird (z007: 29 MUA, 11 HVC_I, 12 $HVC_{X/RA}$; z017: 18 MUA, 4 HVC_I; z020: 19 MUA, 2 HVC_I; z028: 22 MUA, 4 HVC_I, 11 $HVC_{X/RA}$).

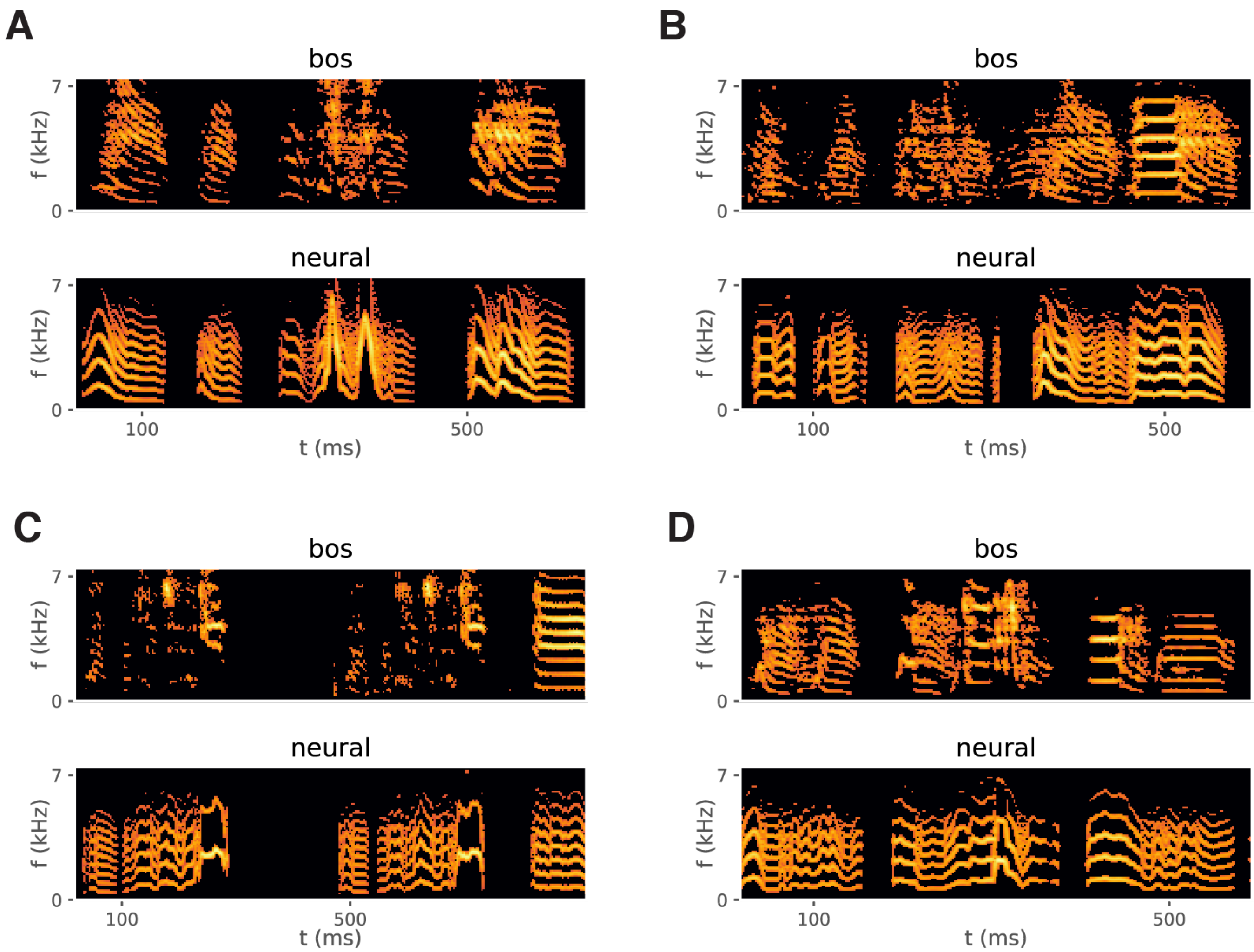

Figure S2. Song synthesized from premotor neural activity via a biomechanical model of the vocal organ using supra-threshold, unsorted spiking events is similar to the recorded bird's own song, Related to Figure 2

(A-D) Spectrogram of a bird's motif (BOS) (upper) and corresponding song generated by inferring the biomechanical model parameters from neural activity using a shallow FFNN and integrating the model, with thresholded activity instead of sorted spikes, for four different birds (z007, z017, z020, z028, respectively; see also Audio S1, S2, S3, S4, respectively).

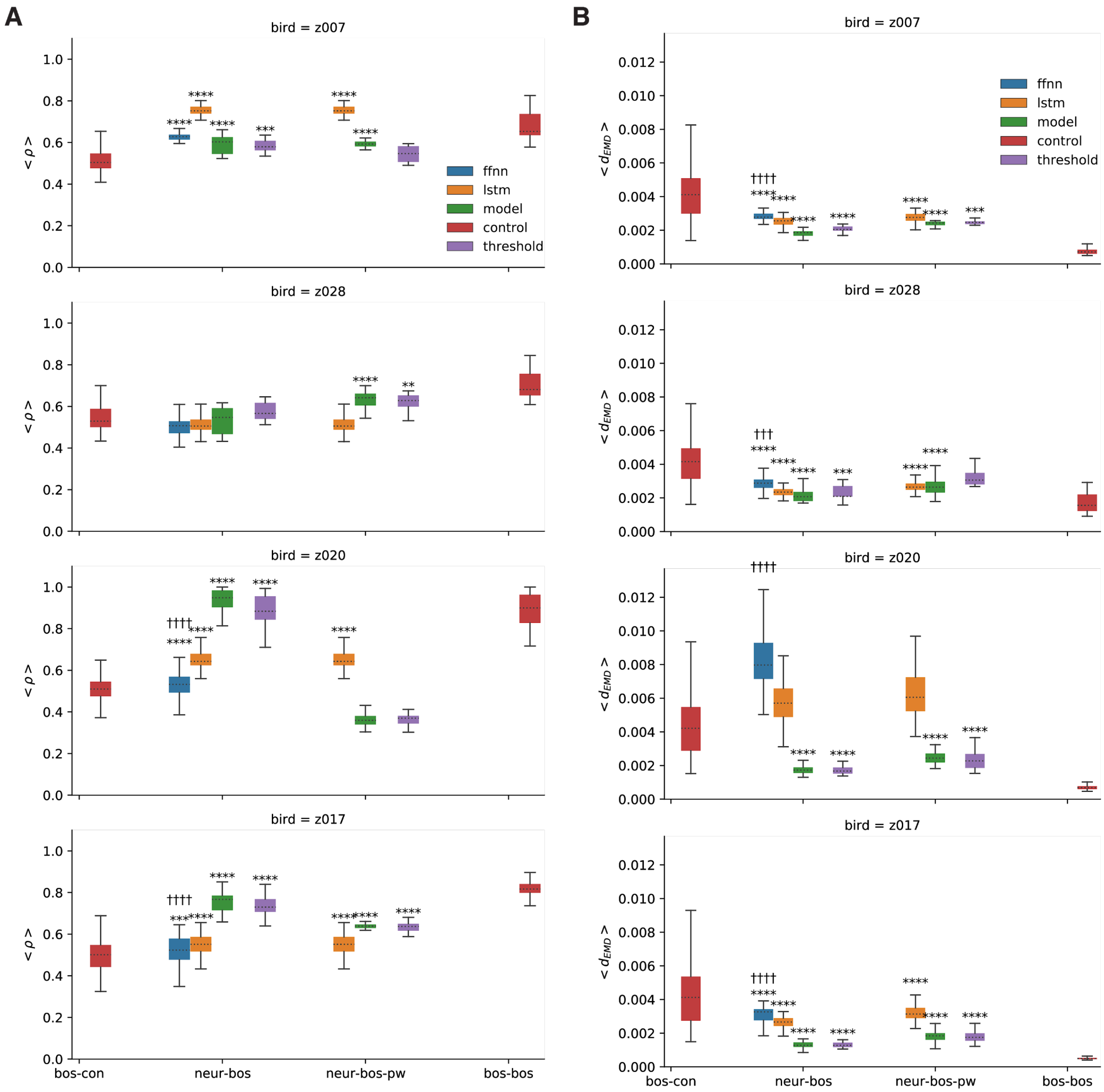

Figure S3. Performance comparisons for each of the birds, related to Figure 4

Legend for each panel is the same as for Figure 4. (A) Pairwise correlations, as in Figure 4 with the addition of piece-wise song training outcomes, for all four birds (z007, z017, z020, z028, from top to down). The termination *-pw* indicates that the training/testing was done piece-wise (see methods). (B) Pairwise EMD for all four birds (z007, z017, z020, z028, from top to down) (**: p<0.01; ***: p<0.001; ****: p<0.0001, Mann–Whitney U test, one sided against *bos-con*. ††††: p<0.0001, Mann-Whitney U test, one sided against (neu-bos, model))

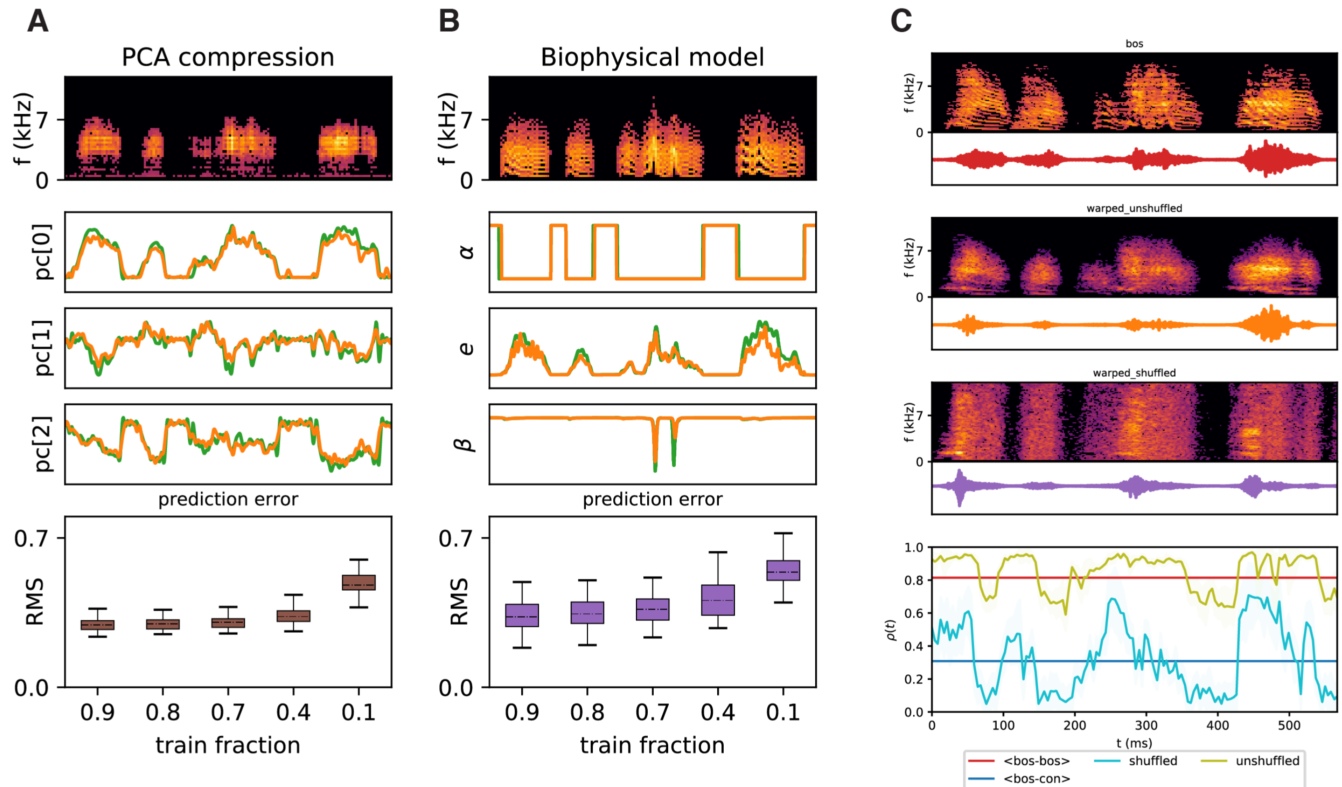

Figure S4. Dimensionality reduction and temporal structure enhance the synthesis of song from neural activity, Related to Figures 2 and 3

(A) Spectrogram of a song motif reconstructed from neural activity after training a FFNN with the 3 principal components (top); detail of the target (green line) and reconstruction (yellow line) of the principal components (middle); error in the reconstruction (as RMS between prediction/target) of the traces compressing the song as a function of the fraction of total motifs in the training set. **(B)** Same as in (A), but with the parameters of the biomechanical model serving as a representation of the song. **(C)** Song synthesized from neural activity with an LSTM, when the temporal structure is destroyed. First panel from top: spectrogram and waveform of a bird's own song (BOS) motif. Second panel: spectrogram and waveform of a reconstructed bird motif, trained and tested on neural and vocal data after DTW. Third panel: spectrogram and waveform of a reconstructed motif, trained and tested on neural activities and warped shuffled vocal data. A shuffle mask was applied to the spectrograms; the shown reconstructed spectrogram has the shuffle mask reversed. Bottom panel: shuffled (cyan) vs unshuffled (gold) performance comparisons, expressed as spectral correlation to BOS across time. Horizontal lines show the mean across time of the bos-bos spectral correlation (<bos-bos>, red), and the mean comparison with motifs from a pool of conspecific birds (<bos-con>, blue).