Independent Manifolds in the Zebra Finch Song: A Strategy for Robust Social Interaction

Daniel Leeds*
Hunter College High School
71 East 94th Street
New York, NY 10128
dleeds@venezia.rockefeller.edu

January 29, 2001

*The author would like to thank Dr. Penio S. Penev of Rockefeller University, for his generous assistance in this research, and Ofer Tchernichovski, for sharing his valuable knowledge and recordings of the Zebra Finch. The author also expresses gratitude toward the Rockefeller University Science Outreach Program for giving him the opportunity to conduct this research and to M. J. Feigenbaum, for generously hosting us in his laboratory, and for his intellectual support and encouragement. Special thanks to Hunter College High School for motivation, support, and the Inter-College Year Program, which provided the author the time necessary to perform research.

Layman's Summary

Zebra Finches are monogamous birds that rely on the song of the male for identification. Separating this song from the auditory rich environment of a bird colony and distinguishing it from those of foreign males is a challenging problem for the female. Here we show that, in order to cope with this problem, the Zebra Finch song has evolved to have complex, nonlinear structure that facilitates both source separation and identification. We develop tools for Independent Manifold Analysis of such structure and apply them both to characterize the song and to reveal the intricate neural control over the nonlinear dynamics of the song-production apparatus.

Abstract

The song of the male Zebra Finch is the primary means of identification of one individual from other males of the same species. In order to serve that purpose, it needs to have a statistical structure that allows robust separation from the rich acoustic environment in the bird colony, as well as enough degrees of freedom to allow specificity at the level of the individual bird. Since second-order structure cannot be used for source separation of linear mixtures—typical in the Zebra Finch acoustic environment—higher-order statistical structures must support the social function of the song. Here we study the dynamics of temporally localized elements of an adult male's song, and find that, when viewed suitably through Poincare sections, the evolution of the sound lies close to a one-dimensional curved manifold, non-trivially embedded in a high-dimensional linear space. On one hand, the nonlinear property of the manifold generally survives linear mixtures, and the one-dimensionality places significant constraints on the structure that could serve as the basis for robust separation. On the other hand, the degrees freedom in the embedding into the high-dimensional space provide enough flexibility for biometric identification of individuals. Also, we discuss how the known properties of the sound-production apparatus of the Zebra Finch can be employed in support of this statistical structure.

1 Introduction

Song recognition is central to the life of the Zebra Finch. Each bird acquires in it's ontogeny a distinct song which is used for identification by other Zebra Finches. Since all Zebra Finches are similarly camouphlaged, song is the sole means by which the birds can distinguish one individual from another. The female finch can only recognize its mate—which it keeps for life—through the song the male sings; ability to distinguish songs is even more important for the female because it lays its eggs in an enclosed nest, from which the female cannot see its mate. Exacerbating the situation, Zebra Finches live in noisy communities in which many Zebra Finches sing simultaneously. Each bird must be able to distinguish the songs of the other birds. Since noisy Zebra Finch communities are natural and can function, it is reasonable to assume the Zebra Finch has developed an efficient sound processing mechanism at the neural level and that the songs themselves, also guided by the brain of the Zebra Finch, have structure, that facilitates source separation and identification. It is well known that source separation of linear mixtures of signals is impossible [4]. Therefore, the bird song must have a higher order structure.

The current method of choice for analysis of the Zebra Finch song is multi-taper spectral estimation [12] followed by the calculation of spectral derivatives, using Fourier Analysis. Though this method has had some success in the study of song learning [11], it has several shortcomings as well. Fourier Analysis does not account for the non-linear structure of the song, analyzing excess noise along with the desired signal. Also, Fourier Analysis assumes a translationally invariant signal, though the Zebra Finch song has several types of well-defined localized structures; Fourier Analysis assumes the signal to be the sum of sine waves—and represents the signal as such—though a quick view of the recording on the millisecond level reveals the presence of distinct repeated waveforms. The Karhunen-Loève Transform [6, 5] (KLT) provides an advantage over Fourier Analysis by representing the signal as the sum of waveforms constructed on the basis of statistics of the signal. Phaseplots of KLT coefficients for an ensemble of like sound activities revealed a low-dimensional manifold for the Zebra Finch song. This manifold was well characterized by Independent Manifold Analysis (IMA), which shows and accounts for the song's non-linear structure.

There is great potential insight to be gained by characterization of Zebra Finch songs by manifold. The changing shape of a sequence type's manifold for a young finch can be analyzed for significance in the field of language acquisition. Further application of IMA on bird songs, and methods to remove sources of extraneous variability, are required, and will be pursued, to fully realize the potential value of the IMA in this area.

2 Timescales of Zebra Finch Song Organization

The song of the adult zebra finch is very repetitious and appears to be very well structured. Fig. 1A shows several " renditions " of the bird's song. Each rendition is composed of multiple " chunks "—

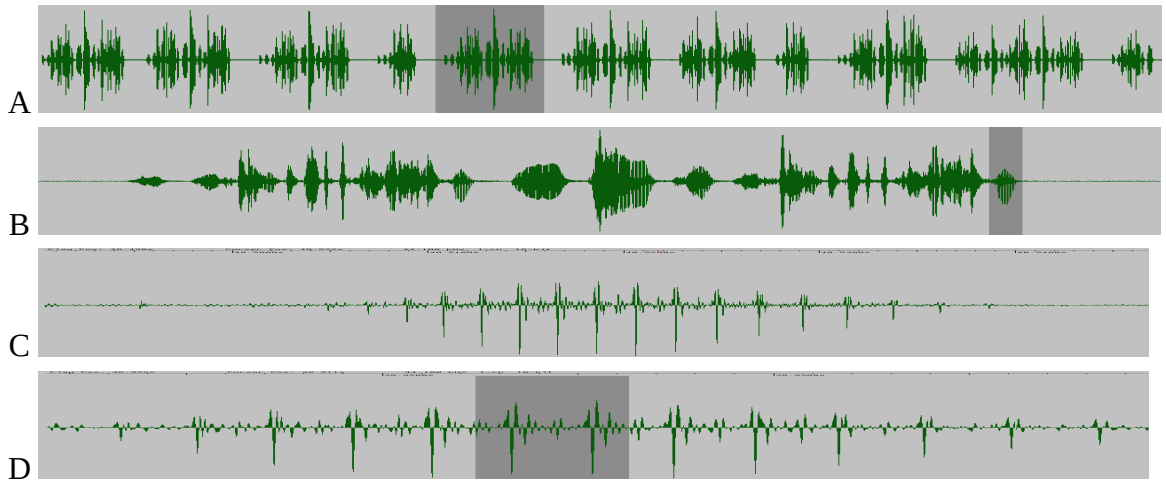

Figure 1: Timescales of song of adult male Zebra Finch recorded in an open cage at 44100 Samples/second. (A) shows several renditions (18500 ms window). (B) displays the interval structure of one rendition (1600 ms window). (C) displays the chunk highlighted in (B) (60 ms window). Chunks of this type are analyzed in this paper. (D) enlargement to 30 ms window size; highlighted area encompasses two pitch periods (3 ms).

sound activities with pauses of 10 to 30 ms on each end. Though not all chunks appear in every rendition, both the " structure " of the chunks and their order appear to be stereotyped (Fig. 1A).

The chunks themselves are relatively well structured. Chunk types can be distinguished in various ways—by duration, by general shape, by amplitude, and by number of syllables, or sounds. However, all chunks have shared properties as well. Each begins with a period of excitation, as its first sound activity begins and increases in amplitude. Each ends with a period of decay, as its final sound activity decreases in amplitude (cf. Fig. 1C).

Each sound in a chunk consists of several repetitions of a waveform (cf. Fig. 1D). This waveform changes in amplitude, shape, and pitch as the chunk evolves. The ability to characterize the changing nature of these waveforms within a chunk would reveal information about the zebra finch's auditory and vocal mechanisms among other things (see Section 7 for detailed discussion). Notably, the chunk and sound activity selected for this research is asymmetric with respect to the time axis. This asymmetry is always realized in the same direction, which suggests intricate neural control over the song production apparatus. Such asymmetry is a robust property of the song, but cannot be captured by multi-taper spectral analysis, which characterizes sound through combinations of sine waves, resulting in an identical representation for the waveform in its recorded state and the waveform if it were flipped with respect to the time axis. In such a case, use of the KLT, which forms localized wave functions for characterization (see Section 3 for a complete description), is clearly beneficial. The selected sound activity type also was interesting due to its structure—unlike several other activity types, the selected type contained both a primary activity

(consisting of high "peaks" repeating approximately every 2 milliseconds) and a secondary activity (consisting of smaller peaks occurring with much greater frequency). The distance between two primary activity peaks is called the "pitch period."

3 The Gaussian Model of Natural Sound

The part of the prepared recording[1] in a given window of duration V samples will be represented by the digitized voltage values $\phi(\mathbf{x})$, where $\{\mathbf{x}\}$ is the sampling grid. An *ensemble* of T waveforms of the sound record will be denoted by $\{\phi^t(\mathbf{x})\}_{t\in T}$. Briefly, [see, *e.g.*, [3] for details] its *KLT representation* is

$$\phi^t(\mathbf{x}) = \sum_{r=1}^{V} a_r^t \sigma_r \psi_r(\mathbf{x}) \tag{1}$$

where $\{\sigma_r^2\}$ is the (non-increasing) *eigenspectrum* of the two correlation matrices

$$\begin{aligned} R(\mathbf{x},\mathbf{y}) &\stackrel{\triangle}{=} \frac{1}{T}\sum_t \phi^t(\mathbf{x})\phi^t(\mathbf{y}) &&= \sum_{r=1}^{V} \psi_r(\mathbf{x})\sigma_r^2\psi_r(\mathbf{y}) \\ C^{tt'} &\stackrel{\triangle}{=} \frac{1}{V}\sum_{\mathbf{x}} \phi^t(\mathbf{x})\phi^{t'}(\mathbf{x}) &&= \sum_{r=1}^{V} a_r^t\sigma_r^2 a_r^{t'} \end{aligned} \tag{2}$$

and $\{\psi_r(\mathbf{x})\}$ and $\{a_r^t\}$ are their respective *orthonormal* eigenvectors. The *average signal power* of the ensemble is

$$\frac{1}{TV}\sum_{\mathbf{x},t} |\phi^t(\mathbf{x})|^2 = \mathrm{tr}\,\mathbf{R} \equiv \mathrm{tr}\,\mathbf{R}_M \stackrel{\triangle}{=} \sum_{r=1}^{V} \sigma_r^2. \tag{3}$$

Eight eigenmodes are displayed in Fig. 2. KLT is optimal in the sense that, among all N-dimensional subspaces ($N < M$), the subset of eigenmodes $\{\psi_r\}_{r=1}^N$ (2) span the subspace which captures the most signal power, $\mathrm{tr}\,\mathbf{R}_N$ [6, 3]. For a given dimensionality N, the *reconstruction* of the snapshot $\phi^t(\mathbf{x})$ is

$$\phi_N^t(\mathbf{x}) \stackrel{\triangle}{=} \sum_{r=1}^{N} a_r^t \sigma_r \psi_r(\mathbf{x}). \tag{4}$$

while *error* is

$$\phi_N^{err} = \phi - \phi_N^{rec}. \tag{5}$$

[1] An adult male zebra finch was recorded from in an isolated, open cage at a 44.1 kHz sampling rate. The audio file and derivative files were held on a Silicon Graphics Indy computer running the Iris operating system. An *ensemble* of $T = 5602$ windows (from 532 chunks) was created. The windows were centered on the peaks of primary activities; these peaks were distinguished from secondary activity peaks via pitch tracking. *Eigenmodes* of $V = 240$ samples—encompassing approximately 2 pitch periods— were formed.

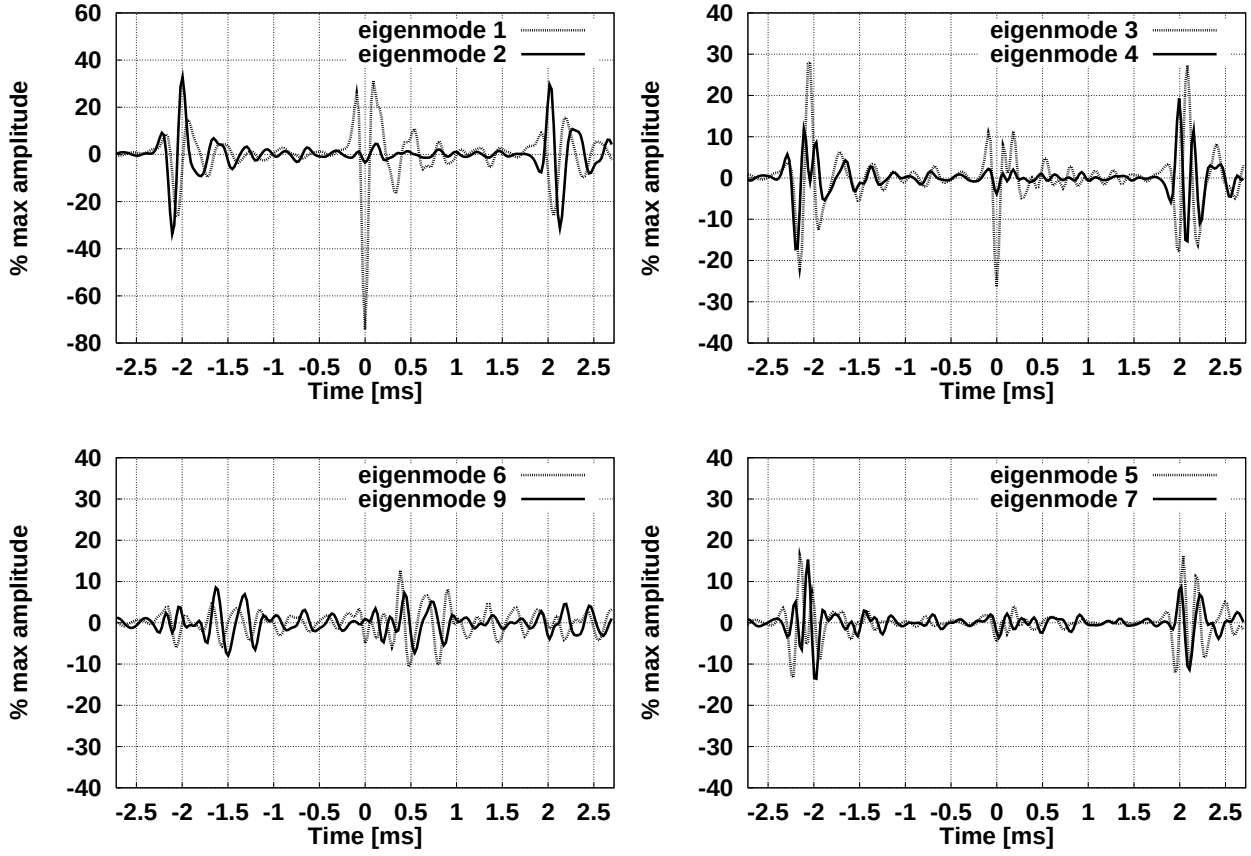

Figure 2: Eight eigenmodes scaled by their respective eigenvalues, $\sigma_r \psi_r(\mathbf{x})$ (1). Note the difference in the scales of amplitudes.

With the standard *multidimensional Gaussian* model for the probability density $\mathcal{P}[\phi]$, the *information content* of the reconstruction (4) is

$$-\log \mathcal{P}[\phi_N^t] \propto \sum_{r=1}^{N} |a_r^t|^2. \tag{6}$$

Notably, this model is *spherical*—the KLT coefficients (1) are of unit variance (2), $\langle a_r^2 \rangle \equiv 1$, and each of the N dimensions contributes equally to the information that is derived from the measurement.

However, as expected, the Gaussian assumption proves to be false. The eigenspectrum (Fig. 3A) reveals three regimes of eigenmodes of decreasing power. The $r \in [1, 90]$ regime is a power-law regime which can be attributed to the structured signal. The $r > 125$ regime is indistinguishable from the random matrices hypothesis [10, 9]. There is a cross-over regime in between.

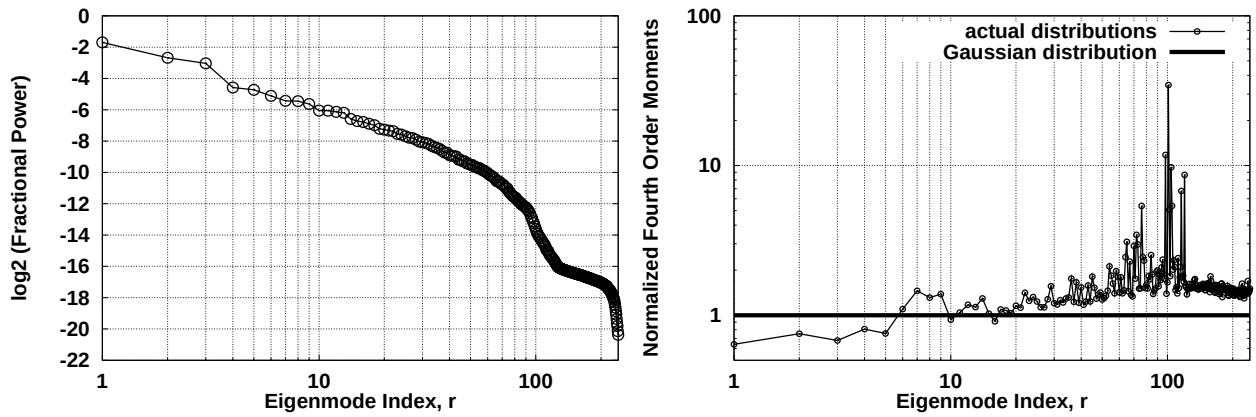

Figure 3: (A) Eigenspectrum of Ensemble. (B) Normalized fourth order moments of the disributions of KLT coefficients, $\langle a_r^4 \rangle / 3$. Notably, $\langle a_r^2 \rangle \equiv 1$. The expectation for a Gaussian distribution, 1, is also shown.

The same three regimes are evident from the normalized fourth order moments (Fig. 3B). As the eigenmodes decrease in power, they rise from slightly sub-Gaussian to slightly super-Gaussian. The greatest points of interest are the sub-Gaussian eigenmodes, the extreme super-Gaussianness of a few of the less powerful eigenmodes, and the phenomenon of the least powerful eigenmodes settling at a level that is slightly super-Gaussian, as opposed to purely Gaussian. The marginal distributions of the first five eigenmodes, which clearly are the most powerful eigenmodes (Fig. 3A), are sub-Gaussian. Comparison of the marginal distributions with the Gaussian model (Fig. 4) further demonstrates this phenomenon, indicating the strong presence of structure in the signal, as expected. The extreme super-Gaussian nature of some eigenmodes (such as $r = 101$) may seem more surprising, but is easily explained: each eigenmode is made super-Gaussian by the random, and infrequent, occurence of a small number of strong instance of the eigenmode. In Fig. 5, the marginal distribution of eigenmode $r = 101$ is predominantly Gaussian, similar to $r = 160$, but there are a few very strong outliers. The presence of moderately strong outliers in the distribution of $r = 160$ explain why it is super-Gaussian. Perfectly Gaussian eigenmodes would be indicated by a concentration of normalized fourth order moments around 1 in Fig. 3B.

4 The Factorial Model of Noisy Trajectories Derived by Vector Quantization

The high level of structure of the first five eigenmodes, implied by Fig. 3B and Fig. 4, indicates correlated joint distributions. The phaseplots reveal a well coordinated motion in a high-dimensional

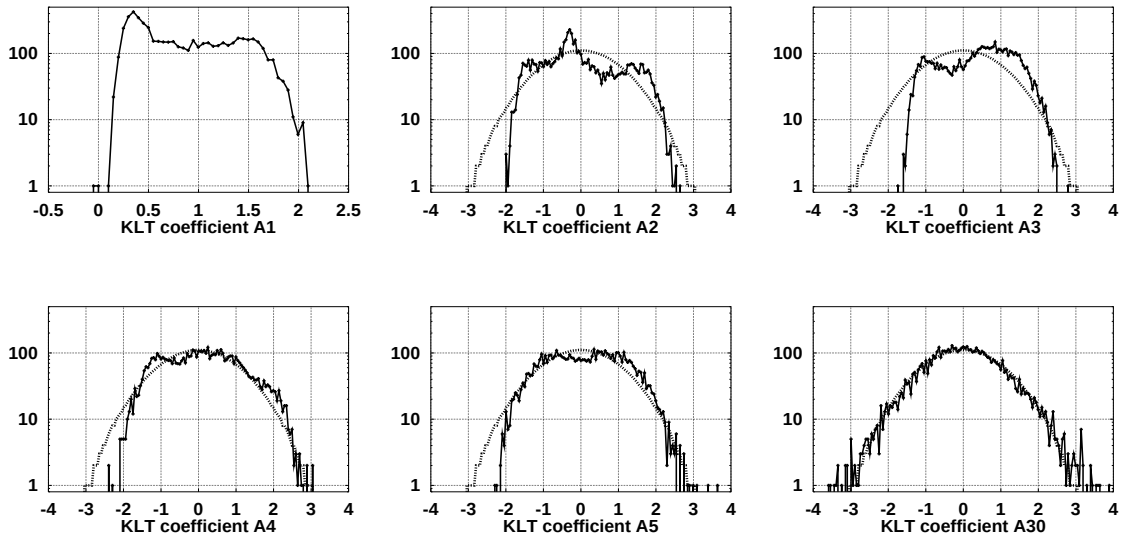

Figure 4: KLT marginal distributions. Histogram counts of the KLT coefficients (1) are shown with solid lines on a logarithmic scale for $r \in \{1, 2, 3, 4, 5, 30\}$, and, with dashed lines, for a Gaussian distribution. The distribution of A1 is asymmetric by design. Nevertheless, it clearly lacks large values.

KLT subspace (Fig. 6). The projections of the trajectory are " fattened " by extraneous sources of variability, such as noise.

Explicit in the formulation of the KLT model (6) is the assumption that the origin has *maximum likelihood*, which is obviously incorrect in the (1:2:3:4:5) KLT subspace. A generalization has been suggested [8] to account for the actual locus of maximum likelihood—modeled as the "midline," $C(\theta, \mathbf{x})$, of the probability density, where θ is a 1-dimensional parameter (cf. Fig. 7). In order to estimate this locus from the data, *vector quantization (VQ)* has been applied previously [7, 8] to tessellate suitably-chosen KLT subspaces into Q *Voronoi cells*, whose centers have the property of *minimum distortion;* in the entropy-based metric (6), this is equivalent to *maximum likelihood.* Further, $C(\theta)$ has been parameterized as a piecewise-linear curve anchored at these centers. Such a model of the (1:2:3:4:5) manifold (cf. Fig. 6) is shown in Fig. 7.

With $C(\theta)$, for any snapshot ϕ^t, a parameter $\tilde{\theta}(t)$, of the dynamics along the manifold, can be found, such that the manifold element $C(\tilde{\theta}(t))$ is closest to ϕ^t according to the metric used for vector quantization; this is a *maximum likelihood estimate.* When the *residual*— the departure from the manifold—is defined as

$$\tilde{\phi}^t(\mathbf{x}) \stackrel{\triangle}{=} \phi^t(\mathbf{x}) - C(\tilde{\theta}(t), \mathbf{x}) \tag{7}$$

the initial probability model (6) can be refined to

$$\mathcal{P}[\phi(\mathbf{x})] \stackrel{\triangle}{=} \mathcal{P}[\tilde{\theta}] \times \mathcal{P}[\tilde{\phi}(\mathbf{x})] \tag{8}$$

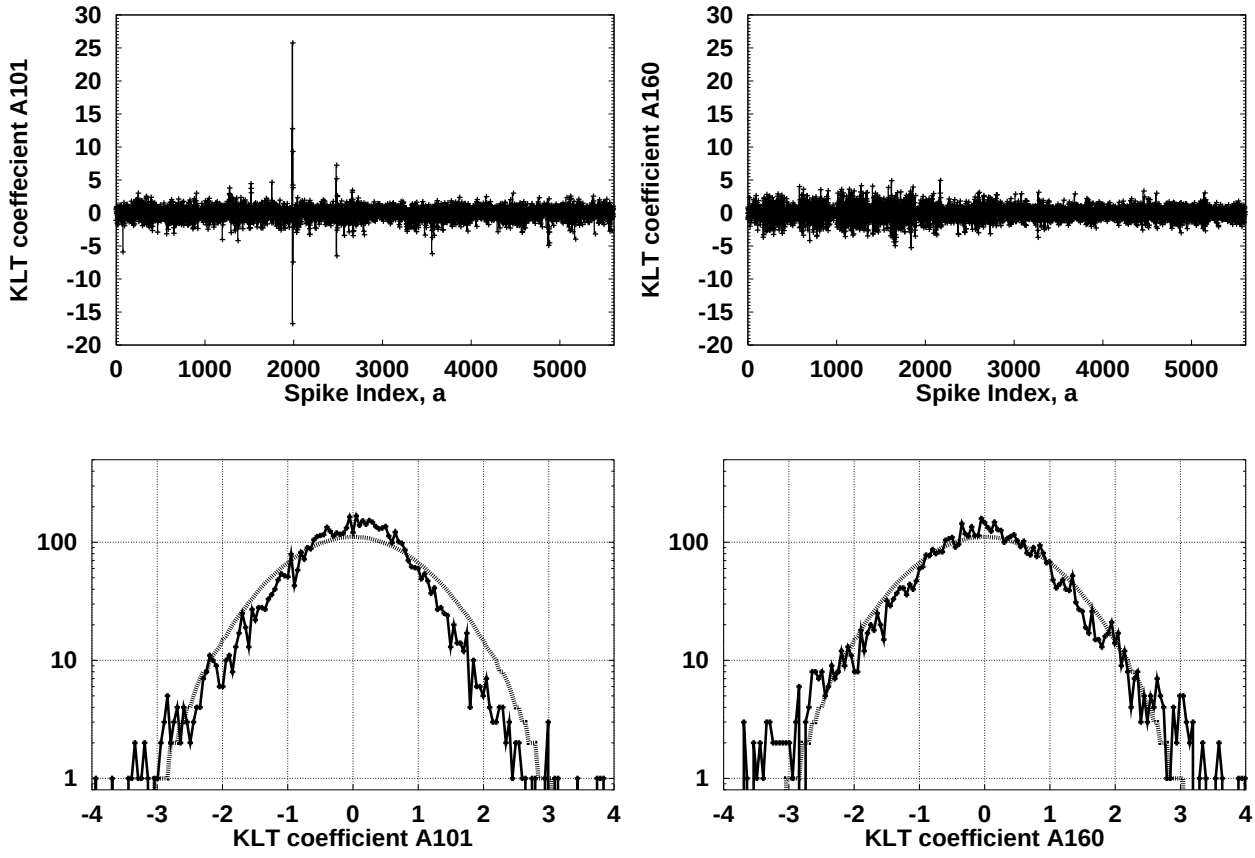

Figure 5: Temporal evolution and marginal histograms of the KLT coefficients of eigenmodes 101 and 160. It is clear that the high fourth order moment of A101 (Fig. 3) is due to a small number of outliers.

where the two factors are estimated independently.

Translating the Voronoi cells of the manifold (Fig. 7A) into wave forms reveals an expected progression. The chosen width of the eigenmodes (cf. Fig. 2)—$V = 240$ samples—allow the eigenmodes to characterize three central, non-stationary traits of the sound activity: primary peak shape, primary peak height, and pitch (inter-peak distance). Focusing on the central peaks in Fig. 8A and Fig. 8B reveals both a progressive change in peak height—from growing to shrinking as the chunk excites and decays. Fig. 8A also shows there to be change in the shape of the wave form portion immediately following the center peak. The changing locations of the peaks in Fig. 8C and Fig. 8D reveals a progressive change in pitch; comparing the activities after the peaks shows a steady change in the shape of the activity follow the peak. There is little change in the middle of the peaks; a set of eigenmodes with lesser volume would form a manifold that better characterized the changing secondary activity.

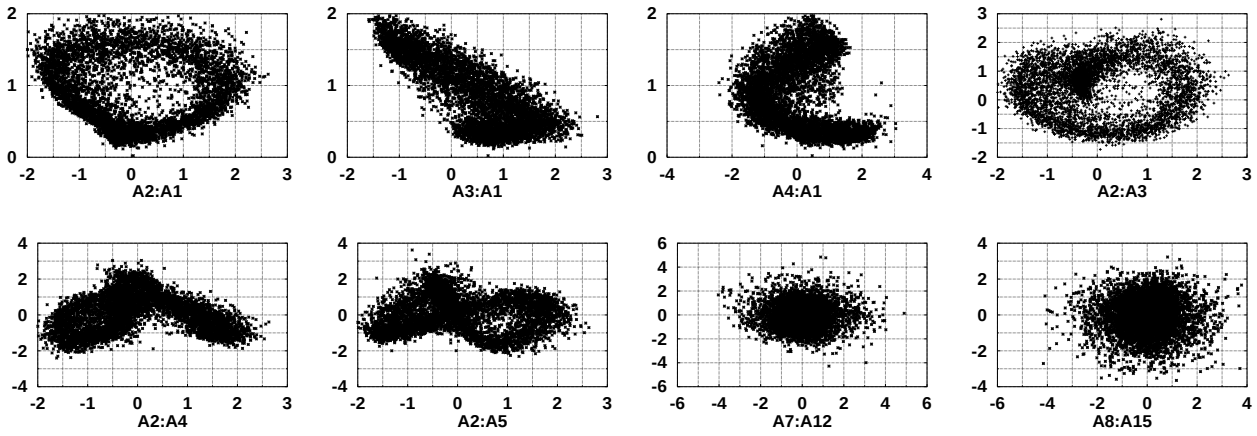

Figure 6: Phase diagrams of the dynamics, $\{a_r^t\}_{t\in T}$ (1), projected on the 2-dimensional KLT subspaces with the indicated pairs of r. The two pairs (7:12) and (8:15), which conform to the Gaussian model, are included for comparison to the structured distributions.

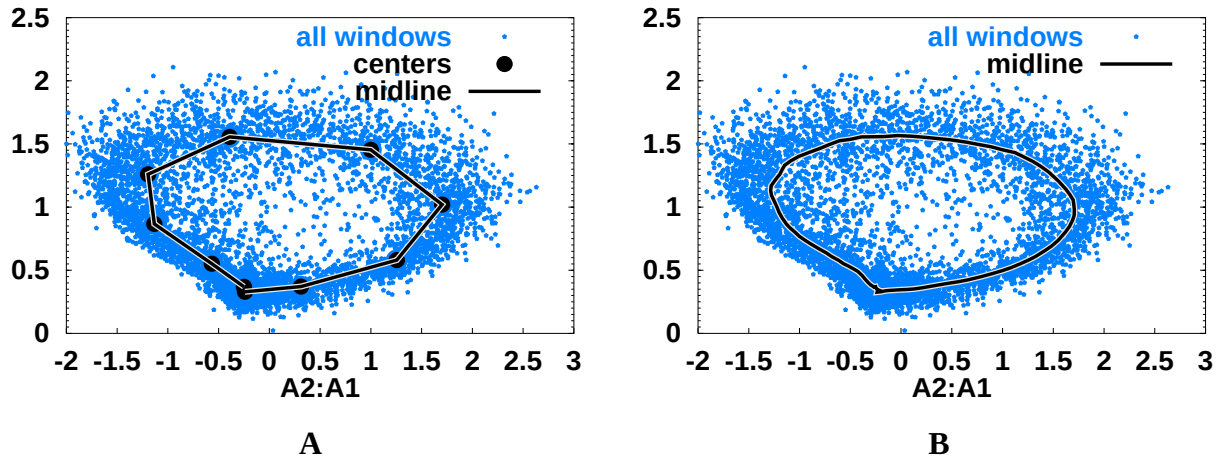

Figure 7: (A) The (2:1) subspace projection of the 1D trajectory, estimated by the LBG vector quantization algorithm, according to the probability metric restricted on the (1:2:3:4:5) KLT subspace. The number of anchors $Q = 10$ was chosen at the end of the 1-dimensional regime of the respective rate-distortion curve. (B) The smooth locus of maximum-likelihood (9).

5 Refinement of the Embedding Parameters

In Section 4, following [8], $C(\theta)$ was modeled as a piecewise-linear curve, in which the anchors were somehow special—they were estimated directly from the data, as averages of the snapshots in a given Voronoi cell Fig. 7*A*. Once an estimate $\tilde{\theta}(t)$ is available though, $C(\theta)$ can be estimated

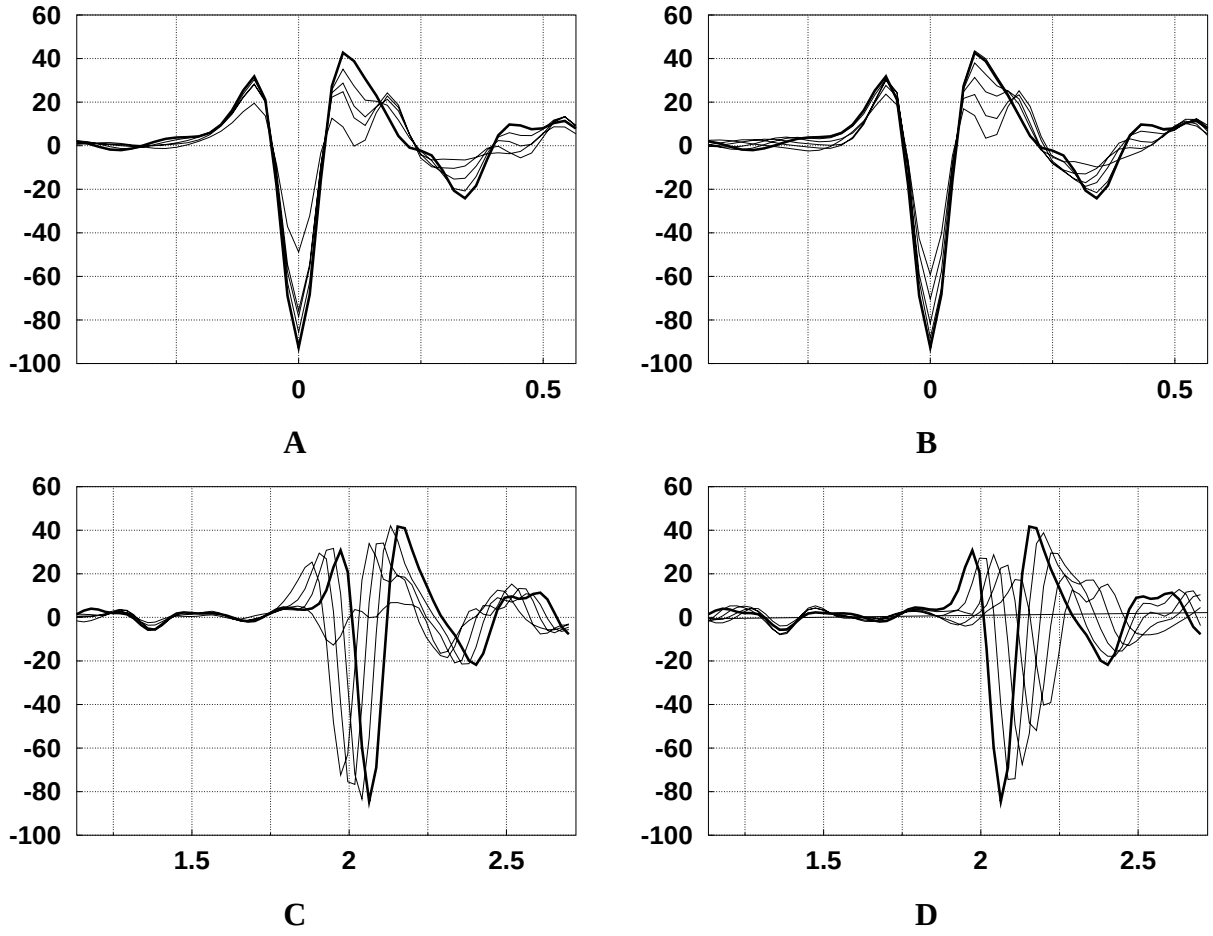

Figure 8: (A) and (B) focus on the center peaks of the Voronoi cells $\theta \in \{4, 5, 6, 7, 8\}$ and $\theta \in \{9, 0, 1, 2, 3, 4\}$ respectively, where the segment of the wave form of $\theta = 4$ is thickest. $\theta = 4$ has the highest peaks and, therefore, is probably the center of the cluster. (C) and (D) focus on the peak to the right of the center peak of each cell's wave form. (C) compares the wave forms before the center of the cluster; (D) compares the waves forms after the cluster center. See Fig. 2 for x- and y-axis units.

for all θ, not just for the anchors, as an average of those snapshots with $\tilde{\theta}(t)$ in some range—of width 2Δ, centered at θ. Then, with $\bar{\theta}(t) = \theta + \bar{h}(t)$, the true parameter on the manifold (which is not known), $\bar{h}(t)$, the true, and $\tilde{h}(t)$, the estimated departure from θ,

$$\tilde{C}_{\Delta}(\theta) \stackrel{\triangle}{=} \langle \phi^t \rangle_{|\tilde{\theta}(t)-\theta| \leq \Delta} = \langle C(\theta + \bar{h}(t)) + \bar{\phi}^t \rangle_{|\tilde{h}(t)| \leq \Delta} \tag{9}$$

where $\Delta = 10\%$ (Fig. 7B) as the number of Voronoi cells used in Fig. 7A was $Q = 10$.

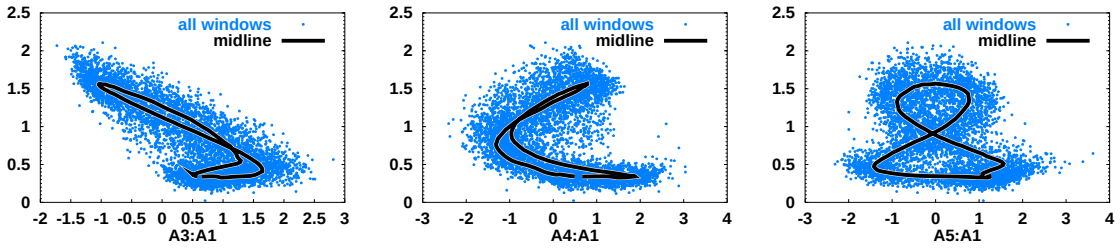

Figure 9: Projections of the manifold, $C(\theta)$, onto the indicated 2D KLT subspaces.

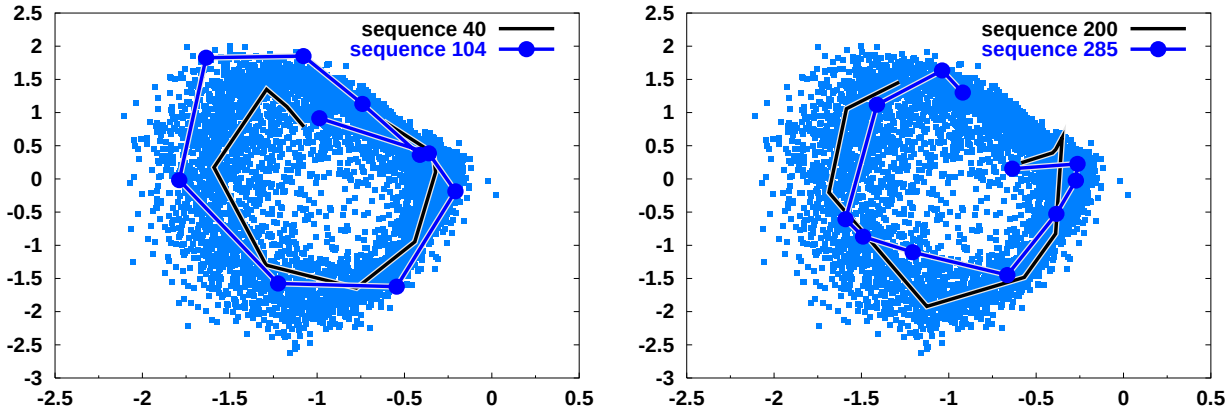

Figure 10: Sequences Following (2:1) Projection of Manifold. Each line connects, in order of occurence, the projections of the wave forms for one chunk, or sequence, in the ensemble onto the (2:1) KLT subspace..

6 Dynamics on the Manifold

Graphing the trajectory of multiple sequences from the ensemble in the KLT subspace reveal a relatively smooth progressions along the manifold formed in Section 4 during the middle of the sequences, but diverse paths used to " get on to " and " get off of " the manifold. Such a pattern is expected, given the means by which the sequences of the song are produced.

The zebra finch syrinx is a nonlinear system, and, therefore, the sound it produces is governed by attractor dynamics, as shown in [1]. Briefly, a vocal system governed by attractor dynamics tenses its muscles to produce a certain sound pattern—the attractor. Regardless of the sound activity the system has been repeating prior to the establishment of the attractor, the system's sound activity approaches and reaches the pattern governed by the attractor. This pattern remains until the attractor is removed. Notably, in such systems mode locking—in which multiple sine waves coincide to form more complex wave functions, the equivalents of eigenmodes—occurs [2, 1].

In Section 2, the excitation and decay structure shown by Fig. 1D indicates the presence of attractor dynamics. Fig. 10 shows that the manifold accurately characterizes the attractor basin phenomenon.

7 Discussion

Analysis of an ensemble of occurences of a certain sound activity from an adult male Zebra Finch's song reveals a high-order structure, as expected. Through use of Poincare sections, one can see that the activity's temporal evolution lies on a 1-dimensional manifold embedded in a high-dimensional space. Views of the trajectories of individual sequences (or " chunks ") in the selected space showed the attractor dynamics characteristic of the nonlinear Zebra Finch syrinx.

Much progress can be made from the work done to demonstrate the non-Gaussian nature of the Zebra Finch song. The ensemble of eigenmodes used did not account for " temporal jitter, " an inevitable inaccuracy in recording. The presence of several derivative eigenmodes strongly indicates room for variability reduction.

The next step would be to study the residual modes. Presence of a manifold for the residual would indicate a high-order structure greater than that established in this paper. Also, further research should be done to characterize the temporal progress of a bird's song along the manifold already formed.

The phaseplots for the eigenmodes of $V = 88$ samples, which represented the ensemble in terms of its secondary activity, indicated the presence of two relatively distinct activities. Further research may reveal a correlation between the location of the activity within the rendition of the bird song and the secondary activity. Also, the presence of two apparently distinct manifolds indicates the need to split the ensemble for more accurate representation of both types of activities. The usefulness of emphasizing the secondary activity may suggest that there are other features of the ensemble that could be emphasized by eigenmodes of a third or fourth volume.

Of course, the work done must be repeated on the other sound activities of the current Zebra Finch song. Each type of activity is likely to have a distinct set of eigenmodes and a distinct manifold. However, comparison of the data for sequences of activities neighboring each other in the rendition may reveal interesting inter-sequence patterns; thinking optimistically, a complex, currently unpredicted, manifold for an entire rendition of the bird song might be possible. Before such information could be revealed, the work must be done.

As previously noted, the sound record used in this study was created by recording a bird in isolation. To prove IMA is robust, experiments must be done on recordings taken in sound rich environments. Though a single male finch's song can be separated from background noise, it is left to be proven that a linear mixture of the songs of many males and females can be separated using IMA, as we have theorized.

Characterization of Zebra Finch song through IMA has important applications in the field of behavioral science. The Zebra Finch, like the human, learns a series of sounds (the equivalent

of a language) through intensive imitation during the early part of its life. It is believed that through study of the young Zebra Finch, one can learn more about human language acquisition. Best understanding of finch language acquisition is gained through observation and analysis of the changes in a young bird's songs. Since a bird song can be represented as a set of manifolds, it can be easily analyzed. Progressive changes in a song over hours, days, and weeks can be represented objectively as changes in the manifolds of the song. Also, by forming songs that travel on or close to the manifold at different speeds, and by slightly altering the manifolds within the identified fields of variability, scientists can determine the degrees of variability of the songs a Zebra Finch can learn. Such conclusions might be extended to human learning.

References

[1] M S Fee, B Shraiman, B Pesaran, and P P Mitra. The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature*, 395(6697):67–71, 1998.

[2] N. H. Fletcher. Mode locking in non-linearly excited inharmonic musical oscillators. *J. Acoust. Soc. Am.*, 64:1566–1569, 1978.

[3] I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York Berlin Heidelberg Tokio, 1986.

[4] Christian Jutten and Anisse Taleb. Source separation: From dusk till dawn. In *Proceedings of the Second International Workshop on Independent Component Analysis and Blind Signal Separation*, pages 15–26, Helsinki, Finland, June 19–22 2000. IEEE Press.

[5] K. Karhunen. Zur Spektraltheorie Stochasticher. *Prozesse Ann. Acad. Sci. Fennicae*, 37, 1946.

[6] M.M. Loève. *Probability Theory*. Van Nostrand, Princeton, N.J., 1955.

[7] Evan Mandel and Penio S. Penev. Facial feature tracking and pose estimation in video sequences by factorial coding of the low-dimensional entropy manifolds due to the partial symmetries of faces. In *Proc. 25th IEEE Int'l Conf. Acoust., Speech, Sig. Proc. (ICASSP2000)*, volume IV, pages 2345–2348, Istanbul, Turkey, 2000. IEEE.

[8] Penio S. Penev, Manuela Gegui, and Ehud Kaplan. Using vector quantization to build nonlinear factorial models of the low-dimensional independent manifolds in optical imaging data. In *Proc. 2000 Int'l Conf. Image Proc. (ICIP-2000)*, pages MA08–11, Vancouver, Canada, 2000. IEEE.

[9] A M Sengupta and Partha P Mitra. Distributions of singular values for some random matrices. *Phys. Rev. E*, 60(3):3389–3392, September 1999.

[10] Jack W Silverstein. Eigenvalues and eigenvectors of large-dimensional sample covariance matrices. *Contemporary Mathematics*, 50:153–159, 1986.

[11] Ofer Tchernichovski, Thierry Lints, Parth P. Mitra, and Fernando Nottebohm. Vocal imitation in zebra finches is inversely related to model abundance. *Proc. Natl. Acad. Sci. USA*, 96(22):12901–12904, August 1999.

[12] Dave J Thomson. Spectrum estimation and harmonic analysis. *Proc. IEEE*, 70:1055–1096, 1982.

[13] R E Zann. *The Zebra Finch: Synthesis of Field and Laboratory Studies*. Oxford University Press, New York, 1996.