

Chapter 1. How isolate songs differ from wild-type songs

Background & Rationale

Like many other oscine songbird species, zebra finches require exposure to adult song to develop a species-typical, wild-type (WT) song. There is a relatively short sensitive period, from about day 30 post-hatch to day 90-100, during which the young male zebra finch gradually learns his song. He does this by constantly changing his vocal output in reference to external (tutor) songs, using auditory feedback. By the end of this period, the young bird's song is structured and stable, and it will remain like that for the whole lifetime of the bird.

In the wild, as well as in various laboratory settings, songs share several species-typical features. When we examine songs developed in a semi-natural colony setting, where young birds are surrounded by siblings, females, and a number of related or possibly unrelated males (all potential tutors), and compare these songs to those developed in a socially impoverished setting, where a juvenile male is housed with a single adult male, we see a similar outcome of song development and of imitation. Even when young birds learn songs from playbacks, in complete social isolation, song development is within the norm, although sometimes less accurate than that observed in live-tutored birds. In contrast to these conditions, zebra finches raised in complete social *and* acoustic isolation are known to produce abnormal or impoverished songs (Thorpe 1958, Marler & Tamura 1964), but very few studies looked at the differences between WT and ISO songs quantitatively.

The earliest study to investigate systematic differences between WT and ISO song was conducted by Price (1979), who raised zebra finches in different social conditions and then recorded their adult vocalizations. To produce ISO songs, he raised young males in either individual or group isolation without exposure to adult song. Group-isolated birds interacted socially and acoustically with their peers. Individually isolated birds only had

acoustic access to other individually isolated birds of similar ages, but a separation wall prevented any social interaction. Although Price couldn't find any indication that they were imitating each other, it seems highly unlikely that the acoustical contact had no influence on their vocal developments and their final songs, because later studies found convergence of song in group-raised zebra finches (Volman 1995). Price's study found that isolates have fewer distinct bout notes, fewer types of call notes, and their note frequencies and durations were higher and more variable. However, the note stability and the general temporal patterning of their song bouts (introductory notes followed by song notes) were quite normal. In addition, long and short calls were less common in songs of isolates.

Thirteen years later, another study undertook a comparison between ISO and WT songs, this time with regards to reproductive success (Williams 1992). The song characteristics were not quantitatively described in this study, and once again, subjective criteria were used to assess whether a song looked ISO-like or WT-like (the authors used the terms untutored-like and tutored-like). Some of these criteria included the prevalence of higher frequency syllables, upsweeps and syllable repetitions in ISO song. Upsweeps and syllable repetitions are virtually never found in WT zebra finch song. ISO songs were found to consist of both ISO-like and WT-like syllables. Overall, ISO males had lower reproductive success, but, strangely, they were more prominent song models to juvenile males. Not only did their sons copy more from them than sons of WT males from their fathers, but unrelated juveniles also imitated their syllables more, and most interestingly, they preferred to imitate the ISO-like syllables and not the normal ones.

The descriptions and analyses by Price and Williams and colleagues were done before more sophisticated and quantitative sound analysis methods were developed, and have limited bearing on the problem of characterizing ISO versus WT songs. Although, there has been some intuitive understanding in the zebra finch field as to what makes ISO song different from normal WT song, we felt that for a clear understanding of the transition from ISO to WT songs, a new descriptive and quantitative study of WT and ISO songs at different timescales of song production is necessary. We set out to fill this gap at first by

examining how spectral song features (e. g., pitch, frequency modulation et cetera) of WT versus ISO songs distribute. We then continued with features that capture longer timescales of song structure – duration of sounds and rhythms. We aimed at developing a metric system that could allow us to assess the “distance” between WT and ISO songs in a general manner, and in such a way that would allow us to test for approximation toward WT songs in subsequent experiments.

Looking at a variety of WT songs, even within the same colony, it is striking how different and diverse WT songs are across birds. In contrast to the intuition one could develop based on the literature, ISO songs are also extremely diverse across birds. Subjectively, some are quite like WT songs, and others are extremely abnormal. We will start by a subjective comparison of ISO and WT songs and will then address the question if, despite the high diversity, ISO and WT songs can be objectively classified into distinct categories, and if yes, in what song features and at what timescales of song structure are ISO and WT songs most separated? Finally, we will examine the cause for the differences between ISO and WT songs, by testing two non-mutually-exclusive hypotheses: that differences are due to the lack of song tutoring and that they are due to social isolation.

Methods

Experimental groups

Wild-type (WT) songs (n=52): To obtain a baseline for WT songs produced by domesticated zebra finches (*Taenyopygia guttata*) in typical aviaries, we used recordings from birds raised in two well-established colonies: The Rockefeller University Field Research Center colony and the Hunter College zebra finch colony. Both colonies have existed for over 20 years. Birds were kept in breeding rooms including family cages (with 6-12 birds) or in larger semi-natural aviaries. All birds were raised in a social environment including the mixed company of males and females. We used 52 songs altogether. We thank Fernando Nottebohm, Noam Leader and Cheryl Harding for making the WT songs accessible to us.

Isolates (n=17): Birds kept visually and acoustically separated from other birds during the sensitive period for song learning are called isolates. We raised 17 zebra finches from the City College of New York breeding colony in this way to create a large sample of ISO songs. All birds were raised by their parents in a dedicated cage until day 7 post-hatch (we start counting from the hatching of the first egg). Zebra finches do not imitate songs heard prior to day 20 post hatch (Roper & Zann 2006). See Appendix I for testing the effect of early exposure on songs. The father was then removed on day 7, and the cage (with the nest box) was taken to a nursery area housing mothers (who do not sing) and chicks only. Birds were raised by their mother, and on day 30, when the young can already feed themselves (and just at the onset of subsong), birds were placed in sound attenuation chambers and kept in complete isolation from day 30 until day 120 post-hatch or later. Colony management and isolation procedures have been described previously (Tchernichovski et al. 1999).

Visualization of songs

To observe the details of the sound frequencies we use spectral derivatives (Tchernichovski et al. 2000). Traditional sonograms represent the power of sound in a time-frequency plan, while the spectral derivatives represent the change of power. For each point of the two-dimensional time-frequency plan of a sonogram, one can measure the change of power from left to right (on time), from bottom to top (on frequency) or at any arbitrary direction. So spectral derivatives are derivatives of the spectrogram in an ‘appropriate’ direction in the time-frequency plan. The derivatives of each point are calculated in an angle that is perpendicular to the direction of frequency modulation. Spectral derivatives can be estimated using multi-taper spectral methods; they have the same resolution as spectrograms and are not artificially broadened.

Sound Analysis

All the analysis was performed using *Matlab 7*, except for feature calculations (below), which were done using *Sound Analysis Pro 2 (SAP II)*.

In our analysis we rely on the calculation of articulation based song features to describe zebra finch songs. Many of the previous attempts to automate the analysis of sound similarity used sound-spectrographic cross-correlation as a way to measure the similarity between syllables. The correlation between the spectrograms of two sounds was examined by sliding one on top of the other and choosing the best match (the correlation peak). However, measures based on the full spectrogram suffer from a fundamental weakness: the high dimensionality of the basic features. For example, cross-correlations between songs can be useful if the song is first partitioned into its notes and if the notes compared are simple. However, even in this case, mismatch of a single feature can reduce the correlation to baseline level. For example, a moderate difference between the fundamental frequencies of two complex sounds that are otherwise very similar would prevent us from overlapping their spectrogram images (a vertical translation will not help since the harmonics will not match).

The cross-correlation approach, as mentioned above, requires, as a first step, that a song be partitioned into its component notes or syllables. This, in itself, can be a problem. Partitioning a song into syllables or notes is relatively straightforward in a species such as the canary in which syllables are always preceded and followed by a silent interval. Partitioning a song into syllables is more difficult in the zebra finch, whose song includes many changes in frequency modulation and in which diverse sounds often follow each other without intervening silent intervals. Thus, the problems of partitioning sounds into their component notes and then dealing with the complex acoustic structure of these notes compound each other. The analytic approach of SAP II addresses both of the above difficulties. It achieves this by reducing complex sounds to an array of simple features and by implementing an algorithm that does not require that a song be partitioned into its component notes.

The analytic framework of SAP II is rooted in a robust spectral analysis technique that is used to identify acoustic features that have good articulatory correlates. The acoustic features that we chose to characterize zebra finch song are represented by a set of simple, one-dimensional measures designed to summarize the multidimensional information present in a spectrogram. A procedure for measuring similarity, based on such an analytic framework has important advantages. First, it enables the examination of one acoustic feature at a time, instead of having to cope with the entire complexity of the song of two birds. A distributed and then integrated assessment of similarity across different features promotes stability of scoring. Second, it is analytic: it evaluates each feature separately and tells you not only that two sounds are similar or different, but also in what sense they are similar/different.

For every 10 ms time window of song, we compute a number of song features to describe spectral changes in song (Fig. 1.1). In this dissertation, we focus on three features: amplitude modulation (AM), frequency modulation (FM) and goodness of pitch. These features have been described previously (Tchernichovski 2000), so we provide only a brief overview here. Amplitude modulation captures changes in the amplitude envelope of sounds. It is positive in the beginning and negative at the end of each sound.

Amplitude modulation is the overall time-derivative power across all frequencies within a range. Units of AM are $1/t$, Sound Analysis Pro 2 does not scale AM, and time units are defined by the 'advance window' parameter. Frequency modulation is the angular component of squared time and frequency derivatives. This measure gives an absolute (unsigned) estimate of frequency modulation. The units are in radians. Goodness of pitch is the peak of the derivative-cepstrum calculated for harmonic pitch. Units are comparable to AM and can be converted to dB by subtracting a baseline and converting to log scale. All the features presented here (except for amplitude modulation) are *amplitude invariant*. That is, the amplitude of the sound recorded does not affect them and hence the distance between the bird and the microphone should have only a minor effect as long as the shape of the sound wave has not been distorted by the environment.

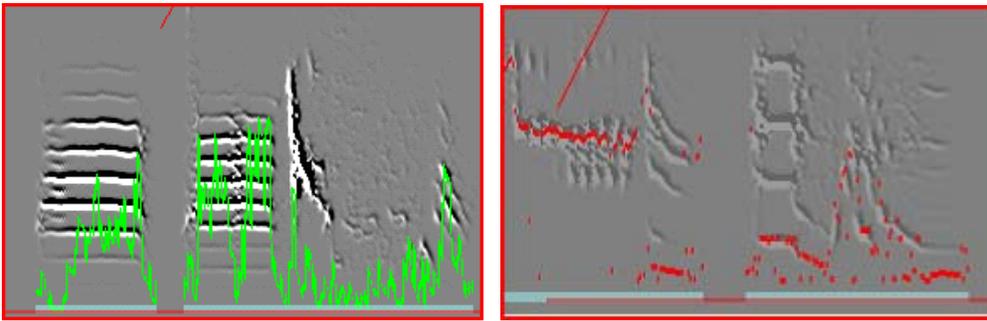


Figure 1.1 | Time series of spectral features overlaid on song sonogram. a. Goodness of pitch (green curve). **b.** Pitch (red curve).

The song features summarize the complex spectral image. A 1 second long sonogram includes 1000 spectral frames, each one represented by, let us say, 500 different frequency values. Therefore, the image includes $1000 \times 500 = 5$ million variables. If we use only 5 features to describe each spectral frame (for example, pitch, frequency modulation, amplitude modulation, Wiener entropy and goodness of pitch), the number of variables decreases to $5 \times 1000 = 5000$. However, this is still a lot, and 1 s is too short a sample to describe the singing behavior of the bird (we usually use at least 20s of song for each bird).

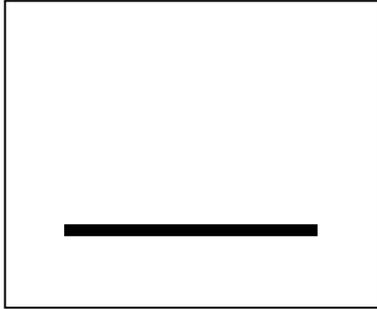
The next stage in simplifying our description is to ignore the time course of each feature (which can be very complicated as seen in the figure above) and look at how the features distribute. The distribution of song features is usually presented as a histogram with bins. The time course of 20,000 pitch values (of our 20 s song sample) can now be reduced to a histogram of 200 frequency values (bins). This decreases the number of variables from 20,000 to 200. However, this is still too many.

The next stage is summarizing the shape of the histogram. In statistics, we often describe distributions using parametric approaches. The most widely used approach is the Gaussian distribution, which is common in nature, and can be fully described by two parameters: mean and variance. This can potentially summarize the 200 frequency values in two measures. However, this approach is only meaningful if the feature distribution is Gaussian, and, unfortunately, this is not the case. Our feature distribution is skewed and often bimodal. In such cases, using mean and variance is rather meaningless – it would tell us very little about the shape of the histogram. Therefore, we are left with the histogram – the so called “non-parametric” description of the distribution (non-parametric is a poor terminology, because what it really means is multi-parametric).

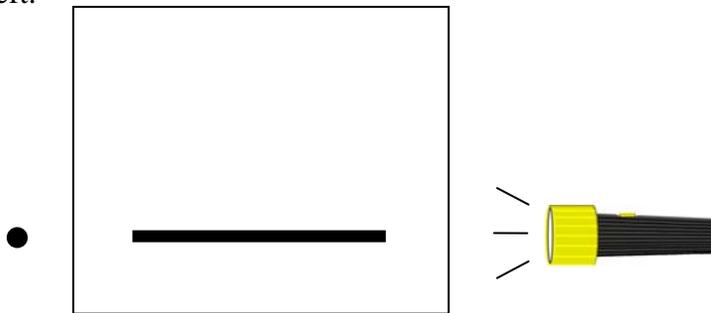
The next question is how to summarize the shape of the histogram without making strong (and often wrong) assumptions about how to model it? For such purposes statisticians developed so called “multivariate statistics” methods, which can be used to compress high-dimensional shapes and images. The most common method is called Principal Components Analysis (PCA).

In general, all the compressions we did so lose information – we cannot go from feature time courses back to the sonograms, nor can we go from the feature histograms back to the time courses. However, we are doing the “compressions” with some belief that the cost, in terms of losing information, is less than the gain: simplicity. Using PCA, we try to simplify our descriptions further at the smallest possible loss of information.

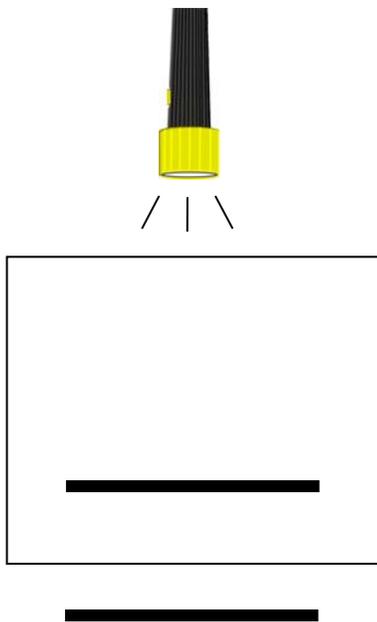
Let us consider a cartoon case – say that we have a piece of paper with a line drawn on it, and our task is to capture this line in a one dimensional representation by projecting it.



We can project the line by placing a flashlight on the right and look at the image on the left.

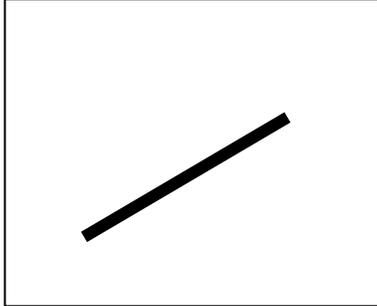


This projection simplified our line greatly, however, we lost most of the information, since the line turned into a single point. The solution is simple: place the light source above the surface of the paper.



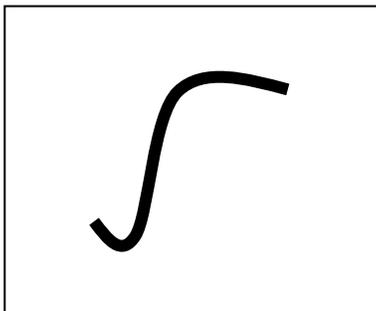
This time we did not lose any information, and we reduced a two-dimensional description (sheet of paper containing the line) into a one-dimensional representation of the object we were interested in describing.

Let us now consider another image.

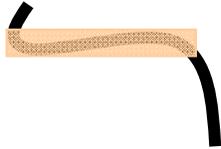


Clearly, we can achieve the same compression as we did with the horizontal line, if we rotate the image in reference to the flashlight. How can we do this automatically? To calculate the exact position automatically, we have to rotate the image slowly and choose the rotation that gives the longest line. That is, the one that maximizes the variance between the points of the line. Mathematically, rotation and projection means that we scale each axis. For example, in the first case of the horizontal line, we simply get rid of the Y axis, namely, multiply X values by 1, and multiply Y values by zero. In the case of a diagonal line, we multiply each axis by a number larger than -1 and smaller than 1. These numbers are called the loading factors or simply coefficients in PCA terminology. They give us all the possible rotations.

Finally, let us consider the following case.



This should be impossible, because there is no way we can capture this shape in one dimension. Nevertheless, some projections reveal more of it and are therefore better than others. We have to find the projection that maximizes the variance. However, there is still lots of information not captured by the projection. We now ask how much of the variance is not yet captured. By removing that “straight line” component of our curve (highlighted), we can now image a much simpler shape, with less structure in it:



Now, we repeat our procedure to capture the most information in the remaining parts of the shape: we rotate to maximize variance and project. Then we repeat again and again. The first rotation is called PC1, the second one PC2 et cetera. When we have rotated enough times, we should have no residual variance left. Each additional PC explains less and less of the variance.

Turning back to our analysis, we need to describe a histogram with 200 bins using PCA. The first rotation is a vector of 200 coefficients. If we multiply each frequency bin value by its coefficient and add them, we end up with one number, this is PC1. Now we go to the second rotation and do the same thing to get PC2.

The critical question is how much we have gained. If most of the variance can be explained by 2 PCs (PC1 and PC2), then we have, in terms of complexity of description, simplified our variables to the same degree as parametric analysis would have. However, in contrast to mean and variance, PC1 and PC2 give an excellent compression without any prior assumptions about the distribution of the data. The more structured the shape is – the better the PCA will work. In an image that is built by white noise and contains no structure, PCA will give no compression at all. So, PC1 and PC2 will explain very little of the variance, and we achieve nothing. PCA works well with high-dimensional data that contains lots of structure but cannot be simply described.

A serious shortcoming of the PCA technique is that interpretation of the results is difficult, because the PCs are just a collection of coefficients that do not mean anything in terms of what the data represents. This is why we find it so important to describe our songs with probability distributions and sonograms first, so we get an intuition about the data. We use PCA to visualize differences between ISO and WT songs and to provide a simple representation that is a good description of the feature distributions on which we can perform statistics.

Results

1.1 The structure of WT zebra finch song

When one listens to isolate zebra finch songs, it is often obvious that these songs differ from the songs of tutored birds. Although ISO songs are extremely variable, they often sound less rhythmic, less structured, more scratchy, longer in duration and higher pitched than WT songs. We will start by describing WT zebra finch songs, and then show common deviations we observe in ISO songs.

Continuous sounds in the zebra finch song are called song syllables. A syllable may be uniform in its spectral structure (Fig. 1.2a), or it may contain several distinct (uniform) sounds, called song notes (Fig 1.2b). Notes are often difficult to identify objectively as they blend into each other and the acoustic transition between them can be gradual rather than abrupt. Although an adult zebra finch song is highly stable, notes sometimes have some variability across renditions.

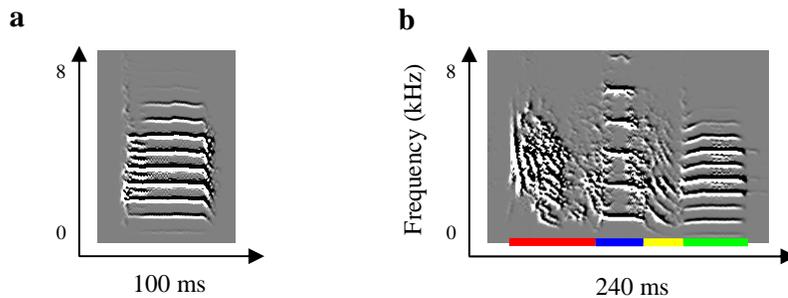


Figure 1.2 | Zebra finch song syllables. a. A syllable consisting of one song note. **b.** A syllable which consists of at least 4 acoustically distinct notes (underlined in different colors).

Syllables usually repeat in a fixed order composing a higher level of song structure called the song motif. A typical motif consists of 2-8 syllables that are separated by silence intervals. Figure 1.3 shows a motif that consists of 3 syllables (underlined in different colors).

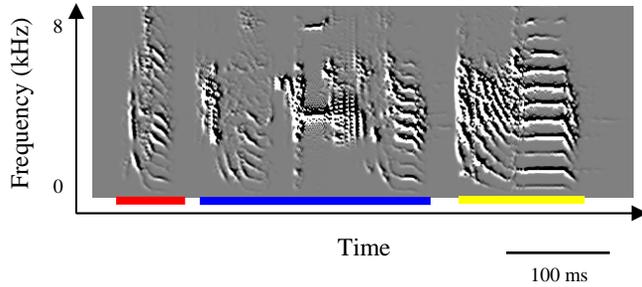


Figure 1.3 | Zebra finch song motif. The underlined syllables make up the motif, the smallest repeating unit of song.

Zebra finches sing in bouts: they first produce a number of introductory notes (Fig. 1.4, underlined in green), which are short, uniform, frequency-modulated calls, and then they repeat their song motifs, usually 2-7 times without stopping. Most zebra finches produce a single, unvaried motif as adults, and their song bouts are highly stereotyped repetitions of these.

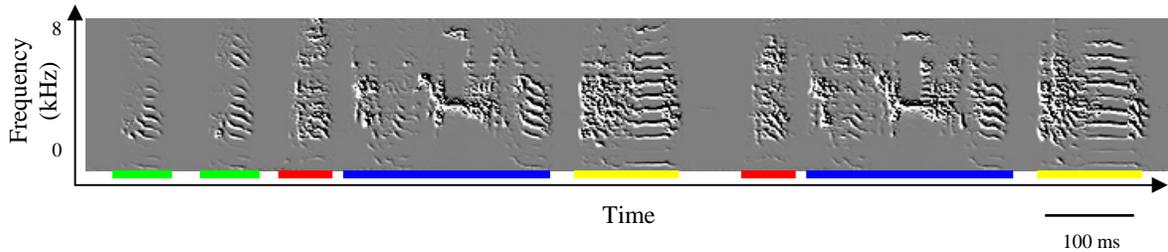


Figure 1.4 | Zebra finch song bout. A bout consists of introductory notes (green) followed by several repetitions of the song motif.

1.2 Subjective comparison of ISO versus WT songs

Some of the diversity of WT zebra finch songs can be seen in the examples presented in Figure 1.5. As shown, the motifs of WT zebra finch songs are highly stereotyped with very little variability between renditions. There are often rapid transitions between notes within syllables, with most notes being short (10-100 ms) and acoustically distinct. Longer syllables are usually complex, contain several notes, but even these syllables rarely exceed 200ms in duration. Most notes have harmonic structure, usually with

distinct pitch (namely, with only one set of harmonic frequencies). Short (Fig. 1.5 Bird 3, green) and medium-length harmonics (Fig. 1.5 Bird 2, yellow) are common with either low (Fig. 1.5 Bird 1, first note of green syllable) or high pitch (Bird 3, near the end of yellow syllable). Many sounds are modulated, with decreasing frequency modulation (down-sweeps, Bird 3, pink) but up-sweeps are extremely rare. Amplitude modulations are strong and fast, and we often see series of extremely rapid amplitude modulations (vibratos, Bird 2 middle of green syllable). Short, high-pitched notes (Bird 4, red) are also common, usually, but not exclusively, within syllables. Long high-pitch notes are very rare.

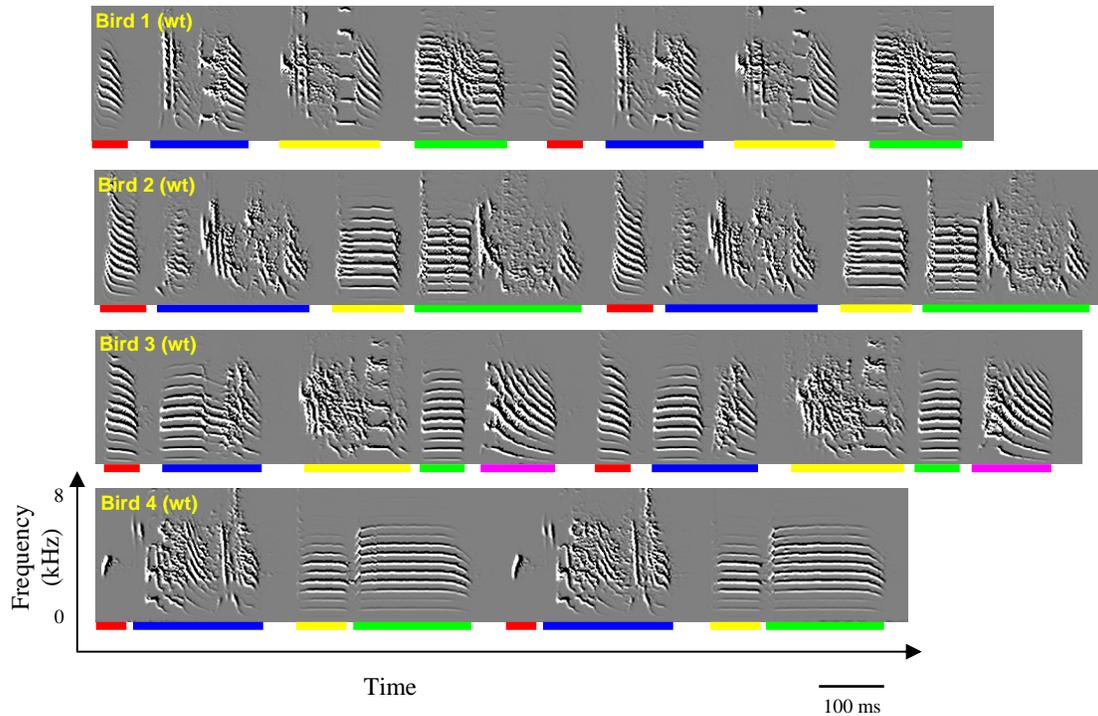


Figure 1.5 | WT zebra finch songs from 4 different birds. Each song example contains two motifs. Syllables are underlined in different colors (red, blue, yellow, green pink in order). WT songs are stable and syllables are always repeated in the same order with very little acoustic variability between renditions. There are fast transitions between notes (Bird 1, yellow syllable), many short and medium-length harmonics (e.g., Bird 4, yellow syllable) and frequency tends to be modulated downwards (e.g., Bird 3, pink syllable).

ISO songs are also diverse across birds, perhaps even more so than WT songs. In Figure 1.6, we present examples that demonstrate the diversity of ISO songs. Some seem as normal as WT songs while others have extremely aberrant song features. In some cases (Fig. 1.6 Bird 1) we see a rather WT-like acoustic structure, but bout organization is unusual. In other cases (Fig. 1.6 Bird 3) the bout structure looks WT-like, but spectral features are abnormal. Bird 1 sings a song that is quite complex with rapid note transitions and normal spectral features, however, there is only one complex syllable that is repeated over and over, and then the bout ends with three harmonic syllables, two of them short and one long. This type of bout organization is unusual and the harmonic syllable is longer than in WT songs. In Bird 2 the song begins with an abnormally long call-like syllable, followed by the stuttering of a single noisy syllable. Bird 3 has a WT-like bout organization, the syllables are repeated sequentially in a stereotypical order, but the syllables themselves are very short and simple, with no or little frequency modulation. Essentially, this bird sings non-modulated harmonic sounds of variable pitches. Bird 4 has a rather normal bout organization, but the most abundant syllable (underlined in blue) is highly atypical, consisting of a high-pitched, long harmonic and a broad-band scratchy sound. Bird 5 sings very simple and long harmonic syllables, and the final syllable (pink) has unstable harmonic structure that makes it sound scratchy. This syllable resembles a juvenile begging call in its spectral structure, and was likely derived from one. Bird 6 & 7 do not have a stable motif and they sing extremely long syllables (Bird 6, green & Bird 7, blue syllables), which are longer than a typical zebra finch motif. The long syllable of Bird 7 is so unusual that even to an untrained ear, it sounds more like a cry of pain than birdsong.

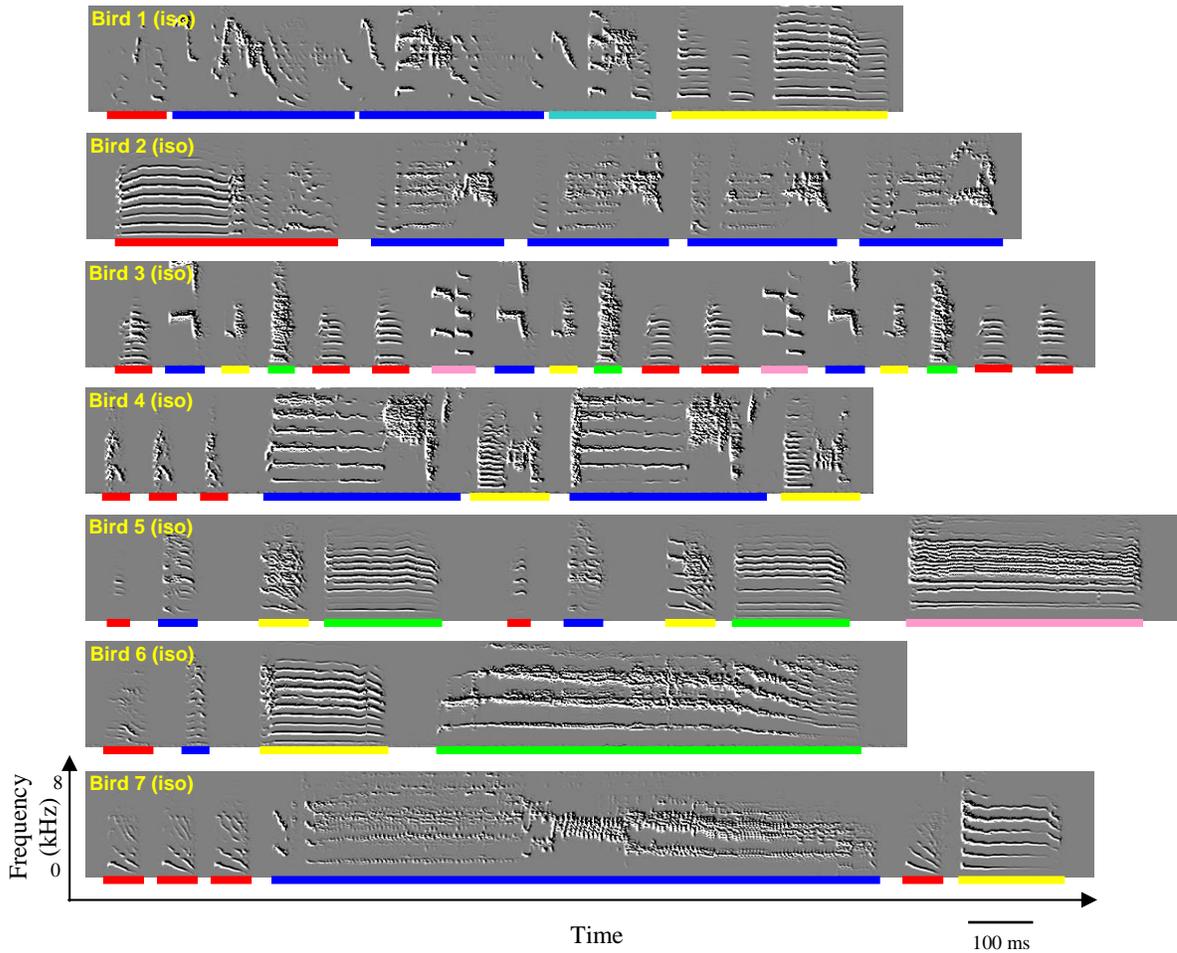


Figure 1.6 | ISO zebra finch songs from 7 different birds. Different colored lines indicate syllables (red, blue, yellow, green and pink in order). ISO songs often have internal repetitions of syllables (Bird 2, blue syllable) but no repeating motifs (an exception here is Bird 3), extremely long syllables (Bird 6, green syllable and Bird 7, blue syllable) and long harmonics (Bird 1, yellow syllable and Bird 5, green syllable).

If we compare typical syllables in ISO and WT songs, we notice that in ISO songs long harmonics dominate whereas in WT songs, short and medium-length harmonics are common. While in WT songs there are many down-modulated syllables, that is, sounds whose frequency contour slopes downwards, these are quite rare in ISO songs. Overall, the spectral complexity of ISO songs is much lower than WT songs. Next, we will

compare song features between ISO and WT songs quantitatively at different timescales and levels of song production.

1.3 Comparing ISO and WT songs with respect to spectral frame features

As we noted above, spectral features such as frequency modulation seem to differ across WT and ISO birds. We tested for differences in the distribution of spectral features by calculating spectral features in 10ms time windows over samples of 20 song bouts per bird (about 10,000-20,000 independent samples per bird). We used *Sound Analysis Pro version 2* to calculate the following spectral features: pitch, frequency modulation (FM), amplitude modulation (AM), Wiener entropy & goodness of pitch. Spectral frame features describe the moment-to-moment vocalization of the bird.

Figure 1.7 presents the mean probability distribution histogram of all our WT (blue, n=52) and ISO (red, n=17) birds for one of the features that showed the strongest differences across WT and ISO: FM. Confidence intervals for the means are indicated by dashed lines. As shown, the distributions of FM show lower values in ISO songs. As we saw in Fig. 1.6, ISO songs contain more harmonic syllables, which developed from long and short calls, and less modulated song syllables. WT songs, on the other hand, have very few low FM sounds, and although in the middle range, there is not much difference, many WT songs have high FM. These are the vibratos and high sloped down-modulated sounds discussed previously.

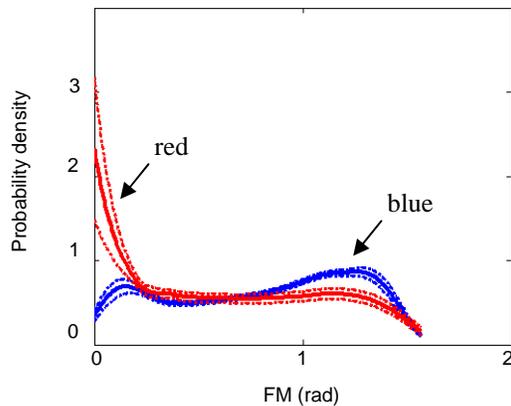


Figure 1.7 | Probability distribution histograms of WT and ISO birds. WT birds (blue line, n=52) show higher FM than ISO birds (red line, n=17). Dotted lines represent 95% confidence intervals.

We found that three features, FM, AM and goodness of pitch, provided the best separation between ISO and WT, so in our further analysis, we will focus on these three features. We calculated cumulative histograms from the probability distribution histograms (Fig. 1.8). The cumulative histograms, of course, showed the same differences between ISO and WT song features that the probability distributions showed, but visually they can show a better separation. ISO FM values are generally lower (Fig. 1.8a), indicating less modulation in sound frequencies. ISO AM values are closer to 0, which means that the changes in amplitude tend to be slower in ISO song (Fig. 1.8b). Goodness of pitch values tend to be higher in ISO song, which indicates that there are more or longer stable harmonic stacks corresponding to long calls (Fig. 1.8c). (See Chapter 1 Methods for information about units of AM and goodness of pitch.)

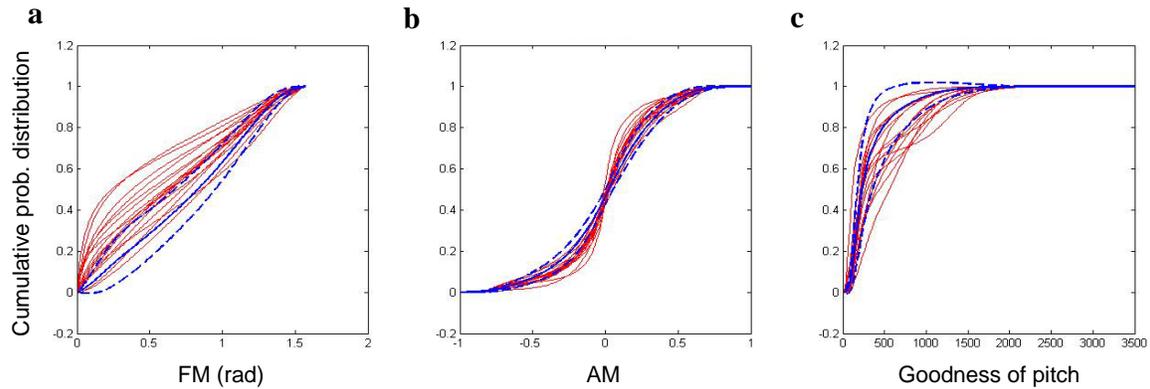


Figure 1.8 | Cumulative distribution histograms of WT and ISO birds for three spectral features: FM (a), AM (b) and goodness of pitch (c). Red lines (n=17) represent individual ISO birds, the blue line is the mean of the WT distribution (n=52), and the dashed lines represent the 95% confidence intervals for the mean of the WT data.

Distribution histograms are very useful to assess the general differences in WT and ISO songs for individual features, but the drawback is that every single feature for every single bird is represented by the long vector of the histogram values. To facilitate comparison across features, we summarized the feature distribution values for individual birds by using Principal Component Analysis (PCA). The input of the PCA is the cumulative histogram vectors of each feature of each bird. The principal components that are returned by PCA summarize the histograms of each individual bird in single numbers.

The first principal component (PC1) combines feature values so as to maximize variance across birds. This is equivalent to rotating the high dimensional histogram spaces of all birds and finding the projection that expresses most of the variance across birds. Then, the residual variance is rotated again to obtain the second PC (PC2) et cetera (more detailed description of PCA in Chapter 1 Methods).

Figure 1.9 shows the result of PCA based on three spectral features: FM, AM and goodness of pitch. Here, the first and second principal components (PC1 and PC2) are plotted against each other. As we can see, the red dots (ISO, $n=17$) and the blue dots (WT, $n=52$) separate from each other, or more precisely, occupy different parts of the continuum. There is some overlap, but almost all red dots are to the right of the blue dots. The distribution histograms of the PC1s of WT and ISO birds are shown under the graph, and they show a clear separation. The mean values of PC1 were significantly different between ISO and WT clusters ($p<0.001$, t-tests, $n_{WT}=52$, $n_{iso}=17$). The PCA does not provide additional information about the specific differences between the songs, but it is an objective means of confirming that there is indeed a real difference between ISO and WT songs.

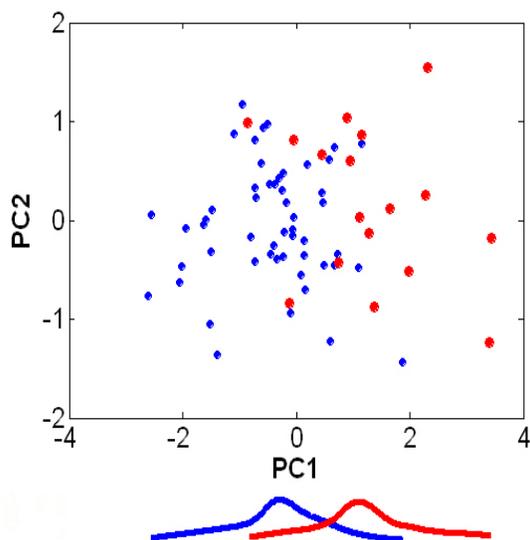


Figure 1.9 | First 2 principle components (PC1 and PC2) of spectral features of WT and ISO songs. Red dots (slightly larger) represent individual ISO birds ($n=17$), blue dots WT birds ($n=52$). Distribution histograms of WT (blue line) and ISO (red line) PC1s are shown under the graph.

We have shown that when all spectral features are taken into account, WT and ISO songs are indeed different.

1.4 Comparing ISO and WT songs with respect to duration of acoustic state

As we mentioned before, ISO song syllables and notes are often prolonged and monotonic. To quantify this notion, we estimated correlation time, namely, the interval over which acoustic features remain highly correlated. Song correlation time can be calculated by computing the spectral auto-correlation of the song bout, and measuring the intervals, starting from the diagonal of the auto-correlation matrix, where the correlation coefficient is higher than a certain threshold. However, during harsh and noisy sounds (e.g., a long monotonic buzz), spectral correlation time is short, even though the bird does not change the features of his singing, so the articulatory state probably remains unchanged. Using features that correlate with the articulatory state provides us with better estimates of correlation time. We therefore calculated the duration of acoustic state based on pitch, FM, Wiener entropy & goodness of pitch. We called this feature the duration of acoustic state (DAS), because it is an indication of the time period that a birds spends singing a sound that in relatively unchanging. We developed an algorithm by modifying an earlier algorithm (Tchernichovski et al 2001) used to calculate the period of repetition in songs.

We scaled the features to units of statistical distances (median absolute deviation from the mean, MAD) based on WT zebra finch feature distribution (Tchernichovski et al. 2000). We then constructed a similarity matrix based on Euclidian distances between the spectral frame features, for each spectral frame in the song. For example, starting with a spectral frame t_i , we calculate the Euclidean distance D from the neighboring spectral frame t_{i+1} using Equation 1.

$$D[t_i, t_{i+1}] = \sqrt{(pitch_i - pitch_{i+1})^2 + (FM_i - FM_{i+1})^2 + (entropy_i - entropy_{i+1})^2 \dots} \quad \text{Eq. (1)}$$

We then calculate the Euclidean distances from the starting point t_i to its more distant neighbors, namely $D[t_i, t_{i+2}]$, $D[t_i, t_{i+3}]$..., and keep going until the distance crosses a threshold of 2.5 Median Absolute Deviations in 3 consecutive spectral frames. This gives

us the right-side duration of acoustic state. We then repeat the same procedure on the left side, namely, $D[t_i, t_{i-1}]$, $D[t_i, t_{i-2}]$.. and combine the left and right side durations to obtain the overall duration of acoustic state (DAS). We consider DAS as undefined when no sound is produced. We implemented this procedure in Sound Analysis Pro 2.

The DAS values for every spectral frame of a zebra finch song motif are shown in Figure 1.10. In this figure, we can see that DAS is an appropriate estimate of the acoustic states of the song, as the values are high for extended periods when the bird produces longer sounds such as harmonic stacks (yellow arrows). In addition, during vibrato sounds, when the frequency modulations are rapid and the momentary spectral features change much, the values largely remain high showing that despite the fluctuations the vibrato state itself lasts long (pink arrow).

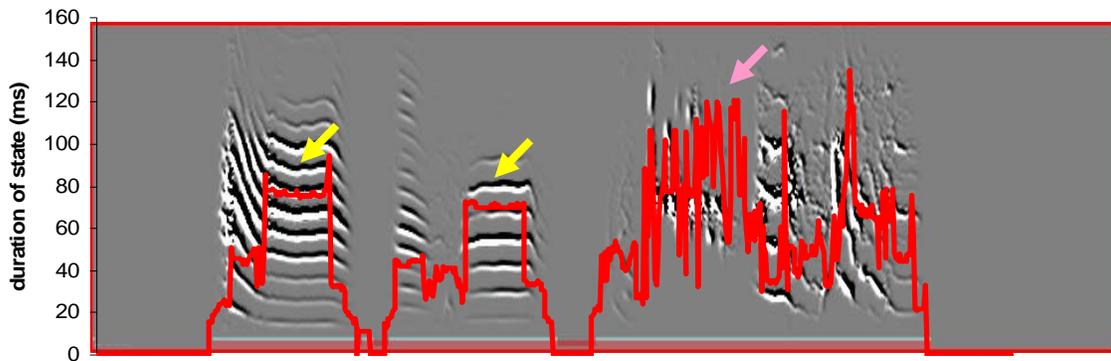


Figure 1.10 | Duration of acoustic state values for a WT zebra finch song. The red line represents the duration of acoustic state values at each time point of the sonogram. Yellow arrows (above first 2 syllables) point at harmonic stacks that have consistently high DAS values. Pink arrow points at a modulated sound (last syllable) where the DAS values remain mostly high indicating that despite the high modulation, the acoustic state is unchanged.

At the beginning of this chapter, we defined a song note as an acoustically uniform unit of song production, so the duration of acoustic state is often similar to note length (Du & Troyer 2006), except that this is a continuous estimate that requires no segmentation of the song.

For every song bout we calculated the values of DAS, which is essentially a long vector just like each of the spectral frame features. We can therefore continue analyzing DAS just as we did with spectral frame features, namely plotting histograms and computing PCA. The probability distribution histograms for the DAS values of ISO and WT birds are shown in Figure 1.11. The histograms show that WT birds (blue line) have shorter durations, in other words, they spend less time in one acoustic state or transition faster between different notes. The ISO curve shows two peaks: the short duration peak is centered on about 40ms, exactly like the WT peak, and the second peak is at about 200ms. The short duration peak is shallower and broader in the ISO songs. The second peak is likely to correspond to the long, unchanging notes often seen in ISO songs (Figure 1.6, Birds 5, 6 & 7). ISO birds may find it more difficult to rapidly switch acoustic states, but whether this is really the case, they show a tendency to get “stuck” singing a particular note.

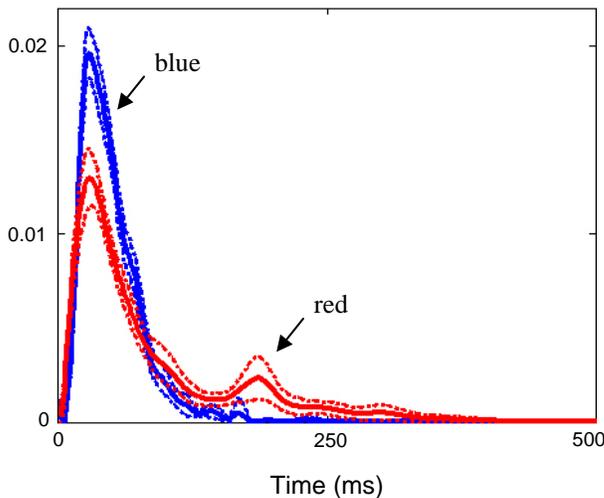


Figure 1.11 | Probability distribution histogram of duration of acoustic state. The red line represents the duration of acoustic state value distributions of ISO birds (n=17), the blue line those of WT birds (n=52). Dashed lines represent 95% confidence intervals.

From the probability distribution histograms, we plotted cumulative histograms (Fig. 1.12) as we had done in our analysis of spectral frame features. For DAS values, the visual advantage of the cumulative histogram over a probability histogram is great as this histogram shows a clear separation between individual ISO birds (red lines) and the confidence interval for the mean of WT birds (dashed blue line). With the exception of perhaps one bird, the red lines (ISO songs) are always below the WT distribution.

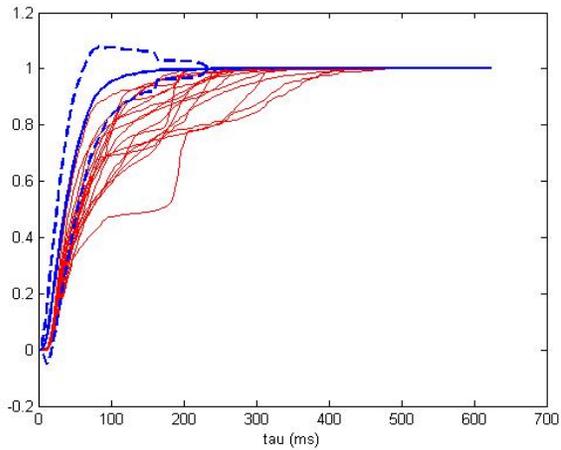


Figure 1.12 | Cumulative Frequency Distributions of acoustic state durations. Each red line represents the song of one ISO bird. The blue line represents the mean values of 52 WT birds, and the dashed line shows 95% confidence intervals.

As with spectral frame features, the DAS histogram is a high-dimensional statistic. To obtain a low dimensional description, we summarized the cumulative DAS distribution vectors using PCA, and plotted the PC1 and PC2 values against each other (Fig. 1.13). Compared to spectral frame features, the separation in the PCA space appears very strong. The inverted U-shape spread of the PCA suggests some correlation between PC1 and PC2, but overall, we see that PC1 separates the WT and ISO distribution much better than PC2, namely all the WT songs are found in the left arm and all the ISO songs are in the right arm. There are only 2-3 birds that are slightly overlapping. The mean values of the first PC were significantly different between ISO and WT clusters ($p < 0.001$, t-tests, $n_{WT}=52$, $n_{iso}=17$).

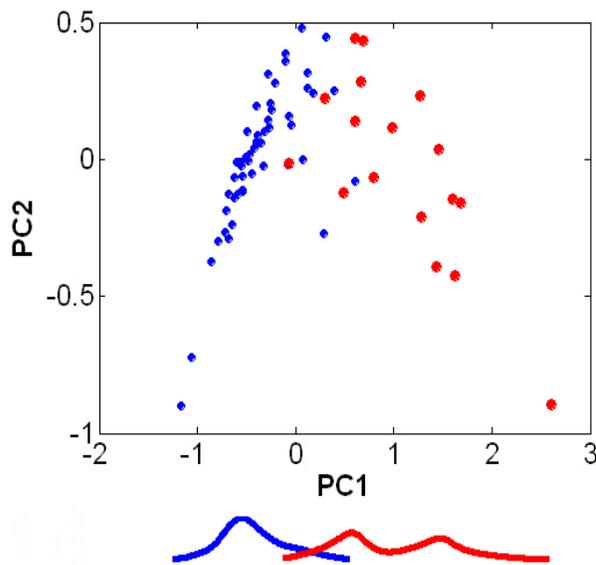


Figure 1.13 | First 2 principal components (PC1 and PC2) of acoustic of state durations of WT and ISO songs. Red (slightly larger) dots represent individual ISO birds ($n=17$), blue dots WT birds ($n=52$). Distribution histograms of WT (blue line) and ISO (red line) PC1s are shown under the graph.

1.5 Comparing ISO and WT songs with respect to song rhythm (bout-level song production)

Although both birdsong and human music are characterized by strong rhythms, there are no standard methods for measuring song rhythm. In western music, rhythm is often explicitly annotated, but the question of how to capture rhythm structure when we hear a natural signal that sounds rhythmic remains open. An existing method measures the distribution of intervals between sounds, and looks for regularities in the sequences of intervals (Tierney et al. 2008), but this approach is ad hoc and has limited generality. Recently, Saar & Mitra (2008) proposed a more objective and general method for capturing song rhythm. Their measure aims at describing regularities in the overall bout organization. Here we used this approach to examine differences in rhythm between ISO and WT birds. We have seen in the examples in Section 1.2 that WT songs have stable repeating motifs, but ISO songs often do not. Instead, there is a tendency to repeat syllables, which are sometimes not fully crystallized. Such differences might be captured by the rhythm of songs, and it is perhaps this difference that is most salient when we listen to WT and ISO songs. WT songs sound very periodic and rhythmic, but most ISO songs do not, or they switch their rhythm mid-song.

The Saar & Mitra method is fairly simple: just as frequency analysis of sound waveform in short windows (of say 10ms) can be used to calculate spectral frame features (pitch, FM, etc.), a similar frequency analysis on much longer windows of singing (e.g., an hour) can capture frequency structure in longer timescales. For example, pitch is the estimate of the period of repetition in very short time scales. For example, when we say that harmonic stack in zebra finch song has a pitch of 1000Hz, what we mean is that the waveform within this sound repeats in periods of 1ms (Fig. 1.14).

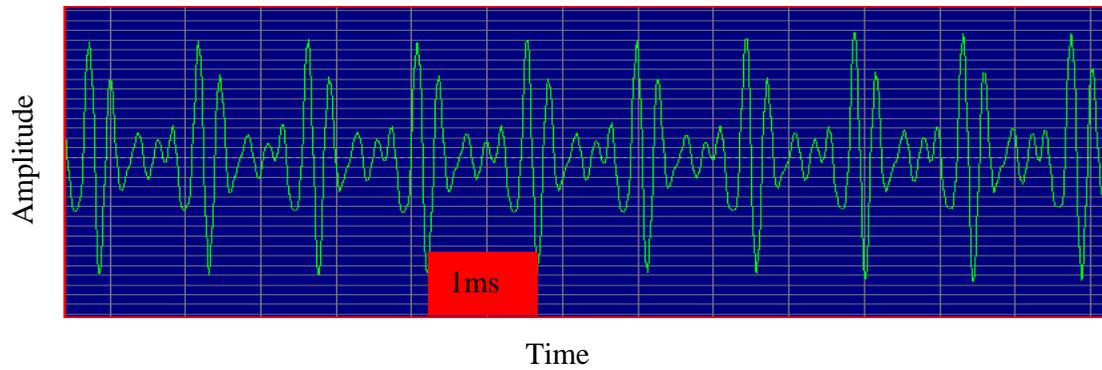


Figure 1.14 | Waveform of periodic sound. Amplitude is plotted against time. Red box indicates the period of the sound.

In a long epoch of singing, when the same motif repeats again and again in a period of 1s, we can also capture the “pitch”, or period of repetition, using frequency analysis in long timescales. Rhythm frequencies can capture patterns of repetitions at both the syllabic level and at the song-motif level. We used a nested spectral analysis method. First, we calculate the spectral frame song features. Then an additional frequency analysis is performed on the song feature time series. The outcome of this second frequency analysis is called *rhythm spectrum*. Although feature values at a given time point depend on the fine temporal structure of the waveform with millisecond resolution, the features themselves change with a slower timescale of 10–100ms. Therefore, the second spectral analysis has most of its energy concentrated on lower frequencies, which allow detection of periodicity in syllable repetition, motif repetition, etc. In the current study, rhythm spectrum was calculated based on the amplitude envelope of song bouts in ISO and WT birds, in the range of 0.5-10Hz.

Figure 1.15 presents the mean and confidence interval for rhythm spectra of WT versus ISO. Note that by averaging spectra across birds we lose most, if not all, information about specific peaks in the rhythm of individual birds, and all we can judge visually is the overall coarse shape of the rhythm spectrum.

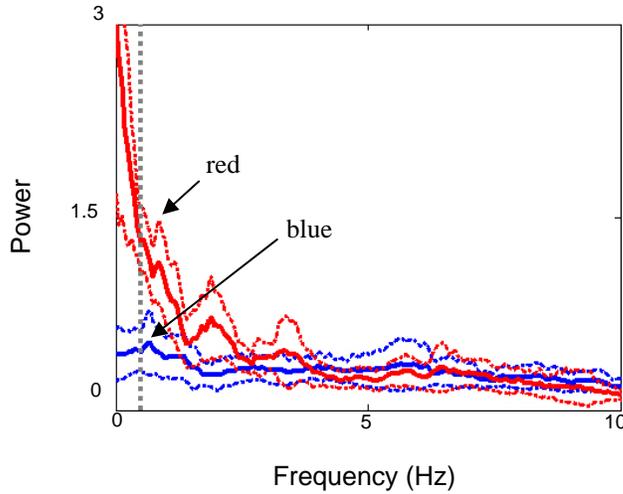
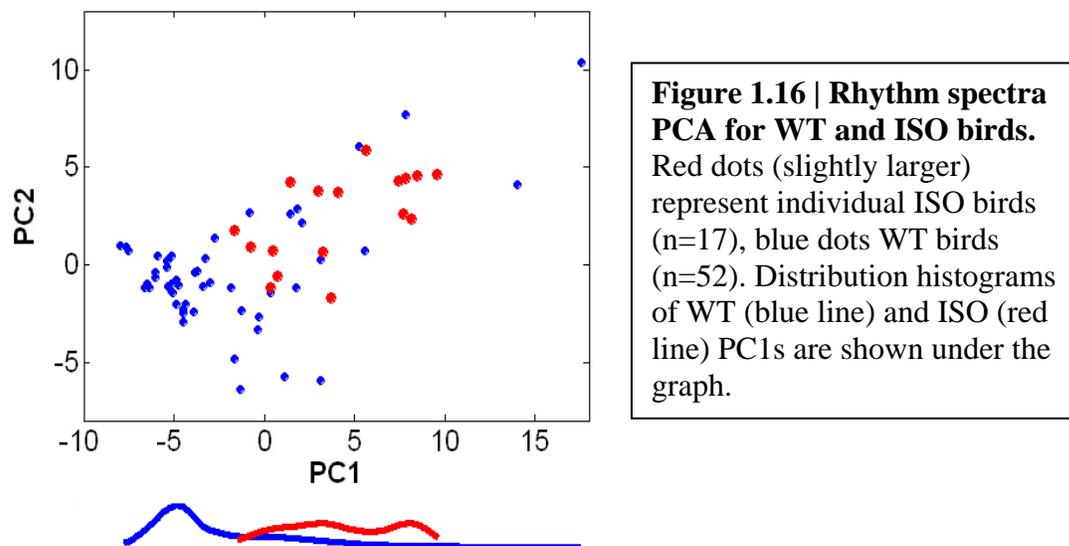


Figure 1.15 | Rhythm spectra for WT and ISO birds. WT (blue, n=52) and ISO (red, n=17) rhythm spectrum distributions. The dashed gray line marks the minimum frequency that we used for further analysis (0.5 Hz).

As shown, ISO songs have much more power in the low frequency range, and the power is going up sharply as we approach zero. There are two possible explanations for this effect: either ISO songs have longer periods of repetitions or the energy we see at the low frequencies of the ISO rhythm spectra is leaking from the zero frequency – which captures the non-periodic energy in the signal. Based on visual inspection of sonograms, which contain lots of shorter rather than longer repetitions (Fig. 1.6 Bird 2) and having listened to many ISO and WT songs, we think that the latter is more likely. Moreover, it could be that ISO song is less rhythmic in the fast repetitions due to jitter in duration of sounds and silence intervals, in which case most of the energy would be concentrated near zero.

As before, we summarize the high dimensional representation (in this case, spectrum) by performing PCA. The results are shown in Figure 1.16. Clearly, the separation in this feature space is not as good as with spectral frame features or duration of acoustic state. Still, the mean values of the first PC were significantly different between ISO and WT clusters for rhythm as well ($p < 0.001$, t-tests, $n_{WT}=52$, $n_{iso}=17$).



In sum, we found significant differences between WT and ISO songs at all timescales of song structure.

1.6 Are the differences due to social isolation or lack of exposure to tutor's song?

Until now, we tested whether ISO songs are different from WT songs. Having confirmed that, we now investigate the cause of this difference. The difference between WT and ISO songs we observed (Figs. 1.5, 1.6) might be caused by the lack of tutoring, but also by social factors unrelated to acoustic experience. For instance, in sedge warblers, deprivation of song tutoring does not result in less structured songs (Leitner et al. 2002). The interpretation of our results would be quite different if the differences between WT and ISO songs using our three song traits (features, duration of acoustic state, rhythm) were mostly due to social isolation, as opposed to the lack of tutoring. To judge this, we used data from birds raised in an arena as shown on Fig. 1.17 (data are from Tchernichovski et al. 2000). We examined if social inhibition of song imitation (Tchernichovski 1998) can lead to ISO-like songs, by allowing 10 pupils to interact with a single tutor simultaneously.

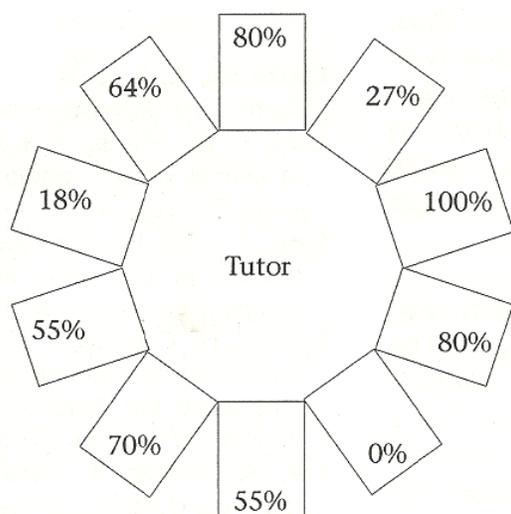


Figure 1.17 | Arena of cages with 10 pupils organized around a single tutor. The tutor was placed at the center, and the 10 pupils around him in a circle. The similarity scores between each pupil and the tutor (manual score) are shown for each pupil.

In such one-to-many arrangements, songs do not converge but diverge. Imitation seems to be inhibited in some pupils but not in others. Given that some birds are more “tutored” than others, we can test if songs that were less influenced by the tutor are more similar to ISO songs. This way, we can test if ISO-like song can emerge in a social situation, therefore excluding the possibility that ISO song is an outcome of social isolation stress. Figure 1.18 presents the similarity to the tutor song against the PC1 of song traits at all three timescales. If lower similarity corresponds to higher PC1 in the ISO range, then we can confirm that ISO song can arise as a result of lower imitation of a tutor and not just in social isolation. The correlation between similarity to tutor’s song and the PC1 is statistically significant for spectral features ($p=0.02$) and for duration of acoustic state ($p<0.01$) but not for rhythm. For the first two traits, the PC values of birds that imitated well project on the WT distribution, and the bird that produced the worst imitation projects onto the ISO distribution. For rhythm, however, all birds appear to be within the WT range even though the trend is similar.

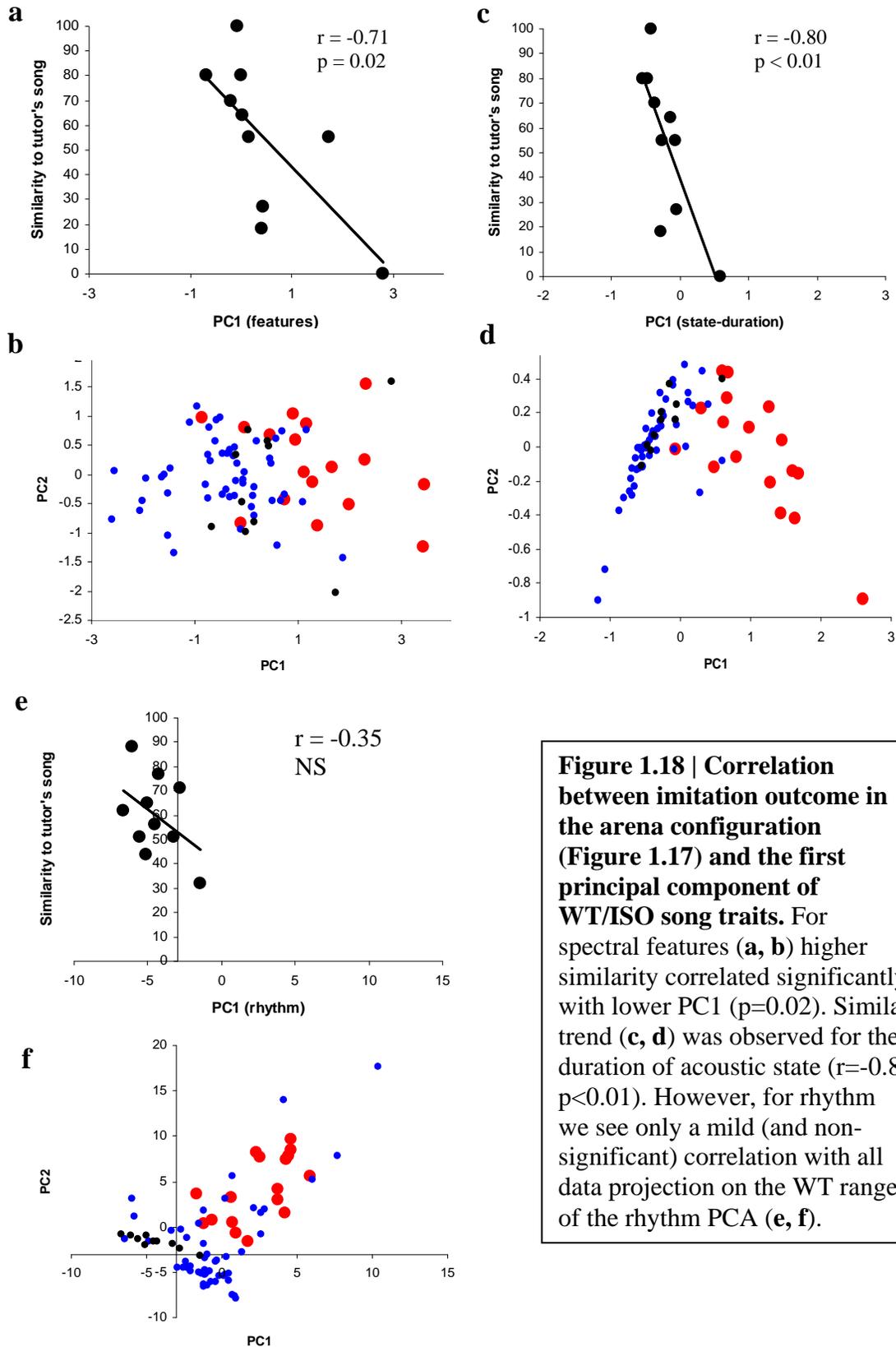


Figure 1.18 | Correlation between imitation outcome in the arena configuration (Figure 1.17) and the first principal component of WT/ISO song traits. For spectral features (a, b) higher similarity correlated significantly with lower PC1 ($p=0.02$). Similar trend (c, d) was observed for the duration of acoustic state ($r=-0.8$, $p<0.01$). However, for rhythm we see only a mild (and non-significant) correlation with all data projection on the WT range of the rhythm PCA (e, f).

In summary, the songs were WT-like in pupils who imitated accurately, but more ISO-like the less the pupil copied from the tutor, confirming that zebra finch ISO songs are, to a large extent, an outcome of tutoring deprivation. There might be an additional effect of isolation stress on zebra finch song development, and such an effect can be tested by raising birds by mute parents.

1.7 Conclusions

In this chapter we showed that when zebra finches are deprived of song tutoring, they improvise a song that is different in many ways from regular WT song. Although there is great diversity in ISO song, deviations from typical WT songs were found at three timescales of song structure: in ms-to-ms spectral features, in the duration of acoustic state and in song rhythm. Differences were significant at all timescales. We will now examine the songs of birds trained by isolates and will use the measures we introduced in this chapter to test if those songs are similar to ISO or WT songs.