Studying the Song Development Process

Rationale and Methods

O. TCHERNICHOVSKI,[a] T. J. LINTS,[a] S. DERÉGNAUCOURT,[a] A. CIMENSER,[b]
AND P.P. MITRA[c]

[a]*Department of Biology, City College of the City University of New York, New York, New York, USA*

[b]*Columbia University, New York, New York, USA*

[c]*Cold Spring Harbor Laboratories, Cold Spring, New York, USA*

ABSTRACT: Current technology makes it possible to measure song development continuously throughout a vocal ontogeny. Here we briefly review some of the problems involved and describe experimental and analytic methods for automatic tracing of vocal changes. These techniques make it possible to characterize the specific methods the bird uses to imitate sounds: an automated song recognition procedure allows continuous song recording, followed by automated sound analysis that partition the song to syllables, extract acoustic features of each syllable, and summarize the entire song development process over time into a single database. The entire song development is then presentable in the form of images or movie clips. These Dynamic Vocal Development (DVD) maps show how each syllable type emerges, and how the bird manipulates syllable features to eventually approximate the model song. Most of the experimental and analytic methods described here have been organized into a software package, which also allows combined neural and sound recording to monitor changes in brain activity as vocal learning occurs. The software is available at http://ofer.sci.ccny.cuny.edu.

KEYWORDS: song development; Dynamic Vocal Development (DVD) maps; sound spectrogram; sound analysis

HISTORICAL PERSPECTIVE: FROM THE SOUND SPECTROGRAM TO AUTOMATED SOUND ANALYSIS

When listening to birdsong, it is immediately apparent that each song has a distinct rhythmic and sometimes even melodic structure. Songs of individual birds in a flock sound similar to each other and differ from those of other flocks. As early as the 18th century, Barrington noted that the songs of cross-fostered birds differed from the species-typical song, suggesting a role for vocal learning.[1] However, until

Address for correspondence: O. Tchernichovski, Department of Biology, City College, CUNY, New York, NY 10031. Voice: 212-650-8540; fax: 212-650-8959.
ofer@ccny.cuny.edu

Software and techniques presented are available at <http://ofer.sci.ccny.cuny.edu>; they can be used freely for basic research on animal communication.

Ann. N.Y. Acad. Sci. 1016: 348–363 (2004). © 2004 New York Academy of Sciences.
doi: 10.1196/annals.1298.031

the late 1950s, there had been no objective way of confirming these observations by physical measurements of the songs themselves. The invention of the sound spectrograph (sonogram) at Bell Laboratories was a significant breakthrough for quantitative investigation of animal vocal behavior.[2] The sonogram transforms a transient stream of sound into a simple static visual image (much like a single frame of film or video) revealing the time-frequency structure of each song syllable. Sonogram images can be measured, analyzed, and compared with one another. This allows the researcher to quantify the degree of similarity between different songs by inspecting (or cross-correlating) sonograms and categorizing song syllables into distinct types. Each song is then treated as a string of symbols, corresponding to syllable types, e.g., a, b, c, d..., and song similarity is estimated by the proportion of shared syllable types across the sonograms of the two songs. The procedure is equally useful in comparing the songs of different birds and that of the same bird at different ages or after control and experimental treatments.

The sonogram has played an essential role in facilitating the mechanistic analysis of birdsong. Much of the pioneering work on song development[3] and on the functional anatomy of the song system[4] relied on sonogram analysis. Indeed, few of the findings in this volume could have been discovered without the invention of the sound spectrogram. Nevertheless, the sonogram has its limitations.

First, the spectral image does not provide simple metrics for characterizing similarity between sounds (except from pure tones). Second, the most prominent features of the sonogram image are not necessarily the most important ones functionally. For example, when observing a spectrogram of human speech, the most apparent features are the harmonic structure and the distinct syllables. However, it is the formants,[d] rather than the harmonic structure, that carry information most important to speech, and it is the language-specific rules, rather than syllable boundaries, that determine the phrasing of words. The raw sound spectrogram is no longer used for analyzing human speech. Instead, modern time-frequency techniques are used to extract perceptually meaningful features that are then further analyzed. In the case of birdsong, however, we usually do not know what the perceptually relevant acoustic features are.

Finally, like the single frame of video, which it resembles, the sonogram presents a static representation of a dynamic process. The sonogram can only capture short-term changes in sound (over time scales of milliseconds), whereas song development is a process occurring over time scales of vocal change ranging from minutes to weeks. The study of that process requires the ability (a) to store large amounts of vocal data, (b) to analyze that data, and (c) to display the results of that analysis as visual representations that highlight key features of the development process as they occur over time. The availability of digital recording and the low cost of digital data storage have now made it possible to meet these requirements.

---

[d] A formant is a peak of frequencies in the spectrum of speech that is caused by the resonance of the vocal track, which enhance some frequencies and dump others. Speech articulation (e.g., moving the tongue or the lips) can change the length of the vocal track, hence, changing the resonance: this is how we produce different vowel sounds (aaa, iii, etc.). (Fucci, D.J. & N.J. Lass. 1999. Fundamentals of Speech Science, 1st edition. Allyn & Bacon. Boston, MA.)

A brief review of some of the techniques developed during the last 20 years to address those limitations is available at <http://ofer.sci.ccny.cuny.edu>, whereas here we present a specific solution that we implemented.

Articulation-based analysis. In human speech, it is possible to extract features of known perceptual function. Such features differ across languages—for example, pitch has phonetic meaning in Chinese, but only prosodic meaning in English. In birdsong, where perception is not well understood, it has been proposed to use articulation-based analysis to extract features with simple relation to production mechanism.[5,6] When considering the motion of an oscillating membrane such as a vocal fold, the most obvious features are the period (pitch) of oscillation and regularity (entropy) of oscillation. These features are also reflected in the sound and can be estimated by calculating pitch and Wiener entropy, respectively. Since sound progresses with time, it also makes sense to measure how pitch changes with time (frequency modulation). Analyzing sounds in terms of such simple features is relatively easy to perform, and the results often suggest an intuitive mechanistic hypothesis (in contrast to spectrographic cross correlation (SCC) and neural-network classification methods).[7–10] In sum, the advantage of our approach is simplicity and interpretability, and its most significant weakness is that the features (and metric system) are "ad hoc," so that some important information about the sound might have been excluded.

AN INTEGRATED SYSTEM FOR STUDYING VOCAL LEARNING

Computation and digital storage cost has decreased tremendously, and with that so have the cost and the efforts involved in collecting and analyzing sound data. A single PC can now handle the recording and the analysis of the entire vocal ontogeny of several birds simultaneously, functioning as a configurable multichannel recorder that recognizes and records songs to digital media and for on-line high-quality analysis of individual songs. However, the design of song storage and analysis systems is a formidable and labor-intensive task, and very few laboratories have been able to invest the time and effort required to design signal analysis tools appropriate to their needs.

There are now several software packages available that can be used for song recognition, sound and brain activity recording, training with operant song playbacks, sound analysis, and song database management. For example, AviSoft (<http://www.avisoft.info/>) is a wonderful integrated recording and analysis application. Other software packages such as Signal (<http://www.engdes.com/>) provide a variety of command-line functions, and Raven (<http://birds.cornell.edu/brp/Raven/>) provides a variety of recording and sound measurement toolboxes. Nevertheless, we found that each software package provides only some of the functions required for managing our vocal learning experiments, and that combining functionality across different software packages is often difficult or impossible (in general, commercial developers do not allow access to their source code). For this reason we decided to develop a new, open code, and noncommercial system designed specifically for the needs of vocal learning experiments—it recognizes and records songs, allows simultaneous recording of neural signal, analyzes features of each sound produced in nearly real-time, trains the birds with operant or passive playbacks, and manages a comprehensive database that allow on-line detection of vocal changes in each bird. This is achieved by constructing descriptive models of song development that show (e.g., as a movie clip) how song features changes, hence complementing the static representations provided by the traditional sonogram with a dynamic representation of vocal output over an extended time scale. Those representations, called Dynamic Vocal Development (DVD) maps stand in relation to the sonogram as a single frame of film does to a motion picture. This system can be easily generalized to support different experimental requirements. The present chapter has two goals. First, we describe the mechanism of automated song recording and sound-analysis technologies that makes it feasible to record the entire vocal ontogeny of an individual bird and to analyze and classify the features of every song syllable produced during song development. Second, because such intensive sampling and analysis of vocal output may seem excessive, we explain the methodological and heuristic advantages that such an approach offers not only to students of song development but for other areas of behavioral neuroscience.

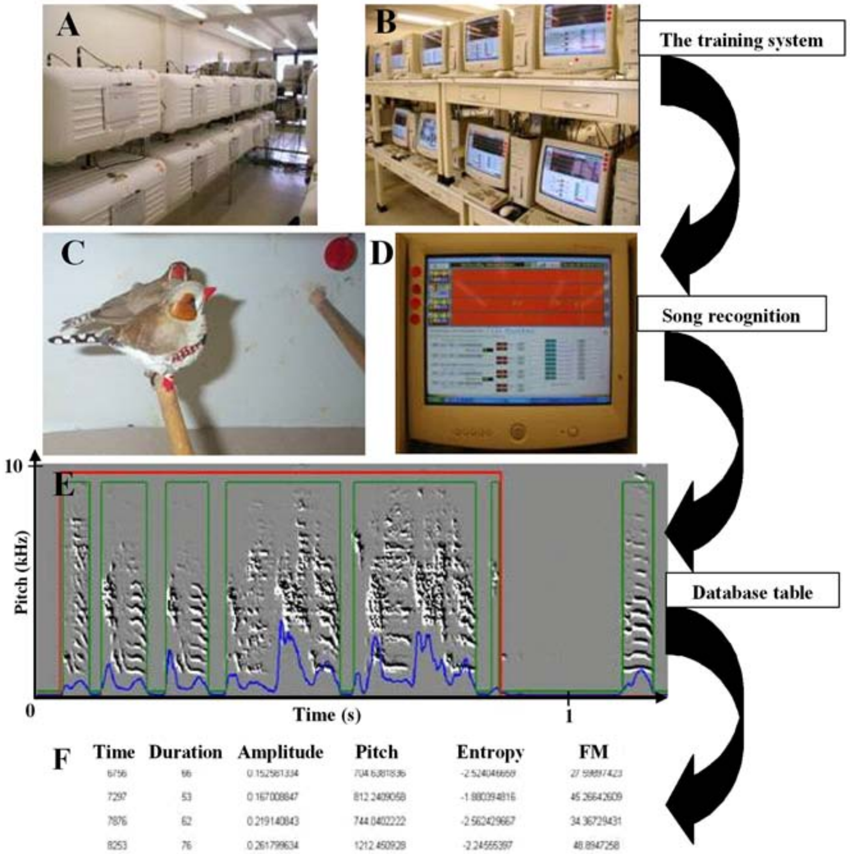
Automated Maintenance of Song Development Experiments

The current version of our recording and analysis system is based upon an experimental unit consisting of a single PC connected to four custom-built soundproof training boxes (FIG. 1A, see <http://ofer.sci.ccny.cuny.edu>). The training regimen for each bird is fully automated and song playbacks are delivered in response to key-pecks (FIG. 1C). Once the bird is placed in the training box, the system records its vocalizations continuously. A song recognition procedure (FIG. 1D) detects and saves the recorded songs, discarding isolated calls and cage noises (for details, see the Sound Analysis user manual at <http://ofer.sci.ccny.cuny.edu>). A few milliseconds later, each recorded song bout is partitioned into syllables (FIG. 1E), then each syllable is analyzed and its time-frequency structure summarized by a set of simple features, such as duration, mean pitch, frequency modulation, etc. These features are promptly saved to a single database file (typically 1–2 million syllables per bird, FIG. 1F).

The Sound Analysis (SA) system (hardware and software) is based on integrating four core functions (FIG. 2): automated multichannel recording, automated operant training regiment, nearly on-line sound analysis, and comprehensive database management. Those core functions are hidden from the user, but they interact with each other at the background—for example, the song detector “knows” when the bird pecked on a key to trigger a playback (and avoid recording it, if so desired).

The function of *training control* is to monitor behaviors and respond appropriately. For example, it responds to a key-peck by delivering a song playback if the training regimen so indicates and registers the key-pecking event to the database. The data card provides many channels and only a few of them are currently in use. The open source-code makes it inexpensive to use those channels as automated on/off switches—for example, to activate a lamp or a buzzer in response to a song syllable, to detect and monitor motion, etc.

Recording control is the most computationally demanding task. It records continuously from four training boxes and, based on song recognition procedures, makes real-time decisions about which recording intervals should be saved for analysis and which should be discarded. The algorithm is based on generic (not bird-specific)



Time	Duration	Amplitude	Pitch	Entropy	FM
6756	66	0.152561334	704.6381836	-2.524046659	27.59897423
7297	53	0.167008847	812.2409058	-1.880394816	45.26642609
7876	62	0.219140843	744.0402222	-2.562429667	34.36729431
8253	76	0.261799634	1212.450928	-2.24565397	48.8947258

FIGURE 1. The training system. (**A**) Training box configuration (40 boxes). (**B**) Training and recording are fully automated by a network of computers. (**C**) A bird and a plastic model in the training box. (**D**) Each computer controls four training boxes and automatically detects singing. (**E**) Spectral derivatives of an early developmental version of a zebra finch song. Spectral derivatives provide a representation of song that is similar but superior to the traditional sound spectrogram. Instead of power spectrum versus time, we present directional derivatives (changes of power) on a gray scale so that the detection of frequency contours is locally optimized. This is particularly useful for the analysis of juvenile song. Songs are automatically analyzed and partitioned to syllables (green outlines) and bouts (red outlines). (**F**) The acoustic features of each syllable are saved in a database table; each row summarizes features of one syllable.

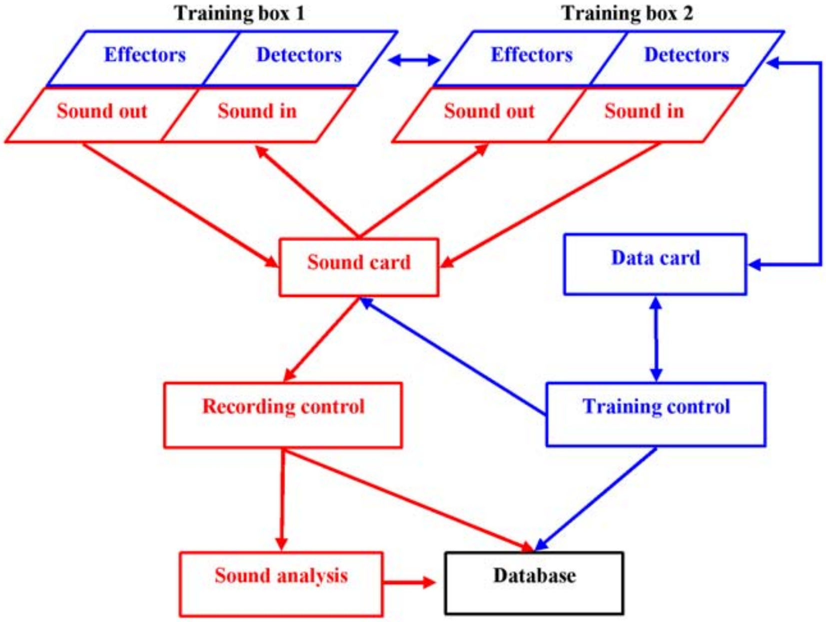


FIGURE 2. A schematic flowchart of the training and recording system.

considerations and works very well for detecting songs during all developmental stages. Each song bout is saved to a separate sound file and attributed with a serial number, bird ID, date, and time-stamps. *Recording control* can also interact with *training control* to cue responses to recorded vocalizations, in nearly real-time, although a highly precise response (such as a delayed auditory feedback) might require additional programming to bypass constraints imposed by the Microsoft Windows® operating system. In principle, a single computer can record and process sound from at least eight sound channels, but we are currently only handling four. Additional channels could be used to control more training boxes or, alternatively, to simultaneously record neural or peripheral data from the same bird. For this purpose, recording channels can be set to trigger each other—for example, a singing event can trigger simultaneous recording from both auditory and neural channels (which are "slaved" to the sound-recording channel).

The function of *Sound Analysis* is to perform on-line measurements to allow song detection and subsequent measurements for parsing the stream of sound to syllables and computing acoustic features of each syllable. The function of the *database* component is to provide easy access to data, including pointers to raw sound data, tables of syllable features, and key-pecking activity, the training protocol and the hatching date of birds, and so forth. We use the mySQL® database server, which is free, open-code application. Several measures, including the amount of daily singing, accumulated key pecking, and DVD-maps (see below) are presented as graphs and are updated in nearly real-time.

Hardware requirements of the system are modest, consisting of a standard PC (~$1,000), a multichannel soundcard (~$400), and an optional data card (~$100). Training boxes are custom designed from coolers and include an airflow system, a microphone, a plastic bird model equipped with a speaker and two keys.[11] A detailed description of how to build the hardware (including the training boxes) and how to use the software is available at <http://ofer.sci.ccny.cuny.edu>. The purpose of the documentation below is to provide a simple description of how hardware and software interact in our system, which is of particular interest to readers who might wish to extend it, e.g., to the processing of neural data.

DYNAMIC VOCAL DEVELOPMENT MAPS

With the entire vocal ontogeny of a bird on file, tracing vocal changes over time becomes straightforward, because we can visualize the raw data of an entire vocal ontogeny in a single image. To demonstrate the power of this approach, we implemented a method suggested by Janata[9] to examine the distribution of syllable durations during song development. An adult zebra finch song is composed of several different syllables; in some cases, each different song syllable has a unique duration. Therefore, plotting a histogram of durations for all syllables produced during a day (about 50,000 syllables) reveals several peaks. Although the histogram is constructed blindly for all sounds produced, it is easy to associate each peak with a specific syllable (FIG. 3A). For instance, syllable 2 is about 140-ms long, so we will consider all sounds around this peak as renditions of syllable 2. How can we know this for sure? We can examine other features such as mean pitch and verify that all sounds that are about 140-ms long have similar pitch, similar frequency modulation, etc. We will elaborate on that shortly, but let us first present a song development image based on syllable duration alone by revealingly plotting duration histograms throughout ontogeny.

Each row in FIGURE 3B presents a daily histogram of syllable durations produced by the same bird, for every day from day 35 until day 90. Overall, this bird produced over a million syllables during its development and all of them are used to construct FIGURE 3B. As shown, every peak presented in FIGURE 3A has turned into a portion of a ridge that can easily be traced back in time. Starting with syllable 1, we can see that the ridge that corresponds to this syllable can be traced back until day 44. Note that training started on day 43, so we can conclude that this syllable type emerged within a single day (this bird was a good learner). The ridge that corresponds to syllable 3 can also be traced and, as shown on about day 55, it takes a turn to the right (when traced bottom-up). This turn indicates a smooth increase in duration, demonstrating a vocal change (time warping) that occurred for this syllable. Finally, examining the ridges from day 35 onward shows that there is no apparent continuity between ridges that appear prior to training and those that appear after training starts.

It is easy to see how additional features, such as mean pitch, can be added to the development map shown in FIGURE 3B, which is the simplest example of a Dynamic Vocal Development (DVD) map. DVD-maps are generated automatically from the syllable database and can be updated in real time, as vocal learning occurs. The representation is robust because it is built on a very large amount of data. Even though parsing the emerging song is not 100% accurate, histograms based on this amount of

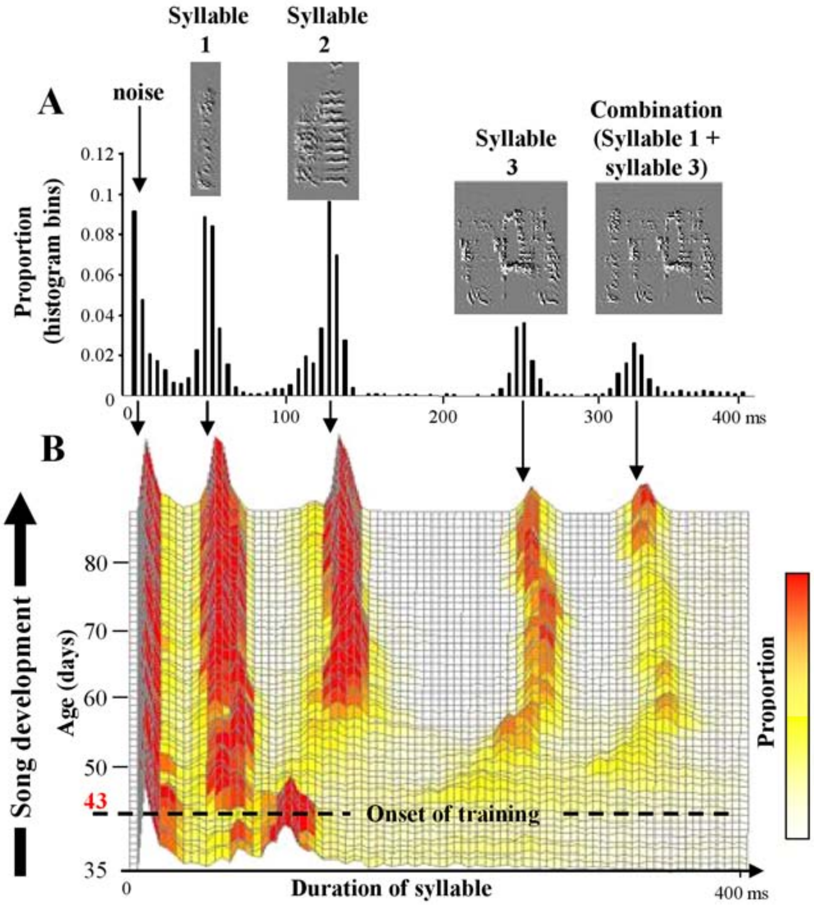


FIGURE 3. Dynamic Vocal Development (DVD) map of syllable durations. (**A**) A histogram of syllable durations in an adult bird. Each peak can be associated with a syllable type of unique duration. (**B**) To trace the evolution of these syllables during song development, we plot histograms of the entire vocal ontogeny. Each row represents the histogram of syllable durations during one developmental day. We can trace each syllable type to an early stage of song development since syllable duration has changed smoothly during song development.

data are insensitive to moderate levels of measurement noise. Inaccuracy in parsing sometimes results in two, partially overlapping versions of syllable boundaries. For example, for a song with three syllables [a, b, c], parsing may occasionally fail to separate syllable a from b, forming joint cluster [a,b] instead of [a] and [b]. Such combined clusters are easy to identify and one can then parse and reanalyze them.

Extending the DVD-map to include more features requires additional dimensions, but FIGURE 3B is already a three-dimensional (3-D) surface. There are several means of adding features to an image without additional spatial dimensions, e.g., using time as a dimension to turn the still image into a movie. To describe how such a movie is constructed, we start with an illustrative two-dimensional (2D) representation of the sounds produced by the same adult male shown in FIGURE 3. Each point displayed in FIGURE 4A plots the duration versus the mean frequency modulation of all syllables produced by the bird during day 90 posthatching. The peaks we observed in the duration histogram shown in FIGURE 3A are presented as *clusters* in FIGURE 4A. Those clusters can be detected automatically by means of cluster analysis (FIG. 4B). Note that the blue and green clusters [shown in color online] are of different frequency modulation (FM) but of similar duration; hence, the ridge in FIGURE 3B that corresponded to syllable 2 was contaminated with another cluster. We can now observe the residuals (FIG. 4C) to see that no additional clusters are available and can examine other features of the clusters (FIG. 4D–G). As shown, we still find the same number of clusters, but on the y axis we can see that the distance between clusters has changed. Some clusters are now more distant from each other (e.g., G

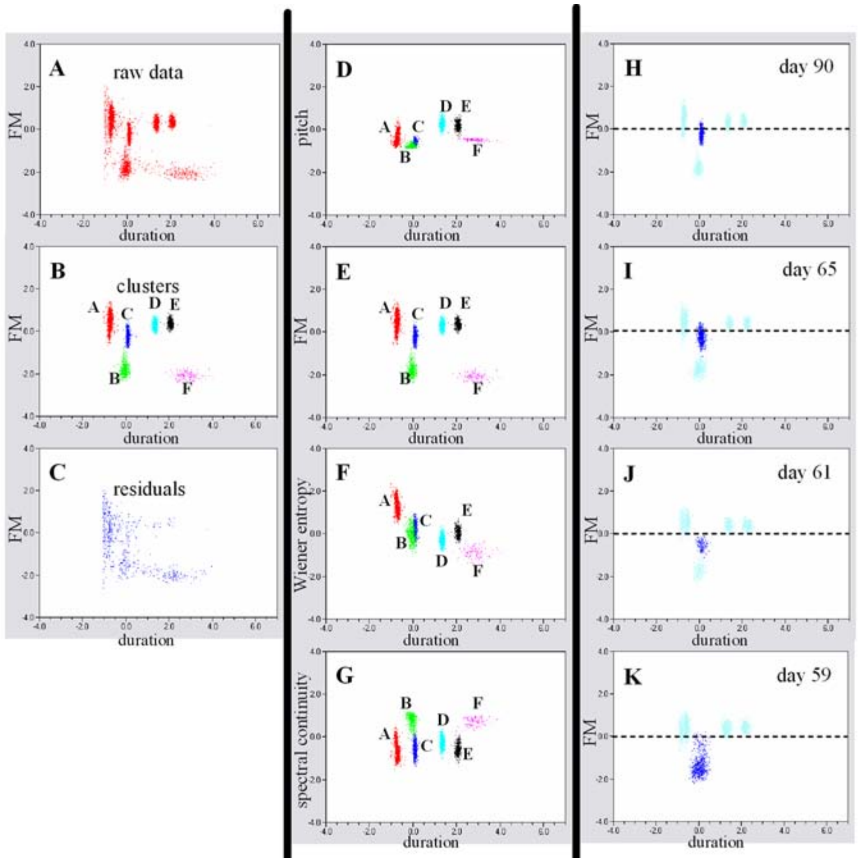


FIGURE 4. (**A**) A plot of duration versus FM, for each syllable produced during day 90 in a bird; (**B**) 10-dimensional cluster analysis of the same data, each color (color shown online) stands for a cluster; (**C**) the leftover residuals of unclustered data; (**D–G**) different projections of the same clusters; (**H–K**) tracking a single cluster.

and F), and other clusters are closer to each other (e.g., B and D). Thus, FIGURE 4D–E can be thought of as different rotations of the high-dimensional image of the clusters. All the projections are consistent with the categorization to six clusters, and it seems reasonable to assume that each cluster represents a distinct type of sound. Nevertheless, the blue and green clusters are close to each other—are they perhaps related? To examine this question, we will have to examine how the image of clusters evolved in time.

The clusters produced from stereotyped adult zebra finch male songs are stably located in these 2D images. Playing a sequence of histogram snapshots, where each

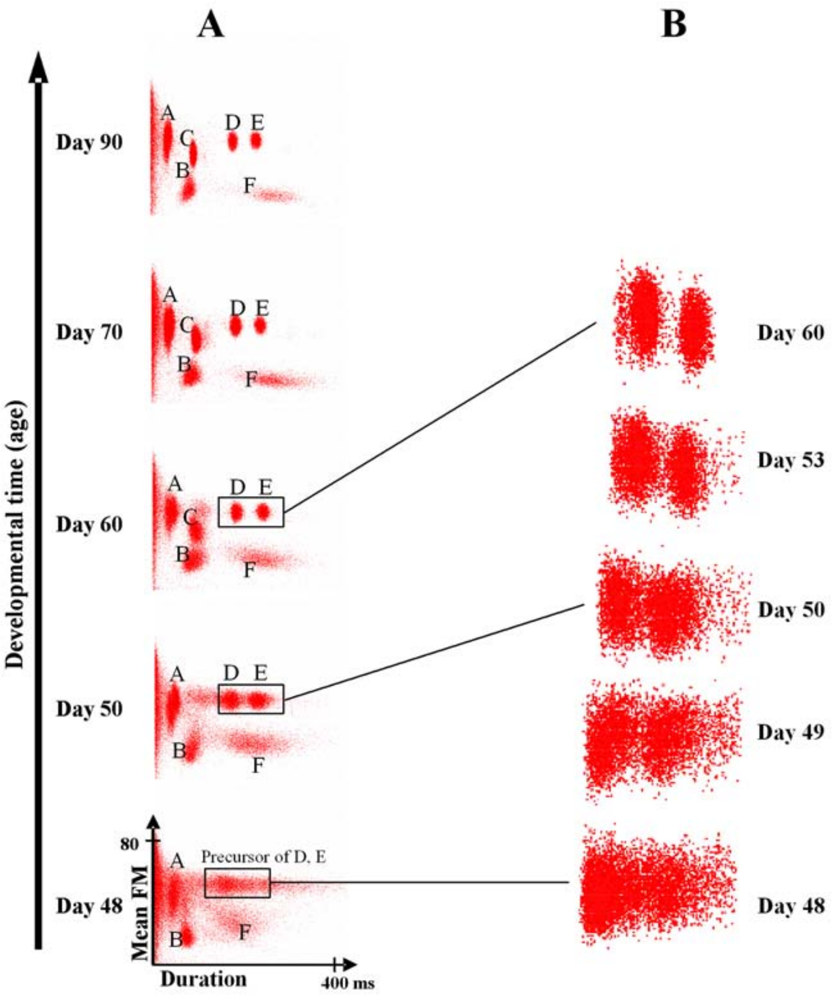


FIGURE 5. Snapshots of DVD-maps (duration, FM) during song development. Six syllable clusters are identified on day 90.

frame corresponds to a narrow time-window, therefore produces a rather static movie clip of sounds produced by the mature bird. When applied to vocal development data, however, such movie clips display very interesting dynamics, revealing the emergence of clusters corresponding to sounds of an imitation. Snapshots of an illustrative movie are presented in FIGURE 4H–K; in those snapshots, the automated cluster analysis routine is attempting to trace the blue cluster only. As shown, the blue and green clusters indeed emerged from the same prototype cluster. FIGURE 5 shows additional snapshots showing how syllables D and E also emerged from the same raw material (the entire movie available at <http://ofer.sci.ccny.cuny.edu>). Finally, we can see the emergence of song-syntax by plotting trajectories that connect one syllable to the next (FIG. 6). In most birds, we observed rapid changes in syntax between days 60 and 70 posthatch.

Overall, clusters emerge, divide, and "move" during song development. Since the features we use are simple, it is easy to interpret such events. For example, by observing DVD-maps of different features, we may detect that a syllable became longer, higher in pitch, or lower in frequency modulation. Each one of these vocal changes can be measured to assess the rate and extent of that change. Based on this type of DVD-map representation, it is possible to examine possible dependencies between different types of vocal change and how many different vocal changes the bird can manage, simultaneously and sequentially, during the imitation process.

DVD-maps represent song development using two complementary approaches: the formal approach involves performing cluster analysis across song development, taking all features into account, whereas the informal—yet equally important approach—is the graphic display in the form of movie clips. None of those approaches would have worked unless we had the entire song development on file. For example, we need about 500 syllables to generate a frame of a DVD-map, but because during each day the bird produces tens of thousands of syllables, each frame is very close in time to the next one. Therefore, each frame is very similar to the former one, obtaining a smooth trajectory of cluster evolution. The DVD movie clips are used to detect vocal changes in real time and to validate, by visual inspection, that the automated procedure did indeed trace the same cluster. A detailed account of this will be presented elsewhere (see <http://ofer.sci.ccny.cuny.edu>), but this chapter will be incomplete without some further elaboration on using a variety of graphic methods to explore changes in syllable morphology and song-syntax development with DVD-maps.

FIGURE 7 [color shown online] presents an example where color has been used to represent the progress of developmental time, with earlier-produced sounds represented by blue dots and later produced sounds by red dots. Over time, the cluster became shorter and higher in pitch. A similar sort of graphic approach can be used to display the history of cluster emergence over longer time frames by using colors to represent different time scales. For example, in FIGURE 8A [color shown online], the feature values of previously produced sounds are plotted in yellow, providing a long-term memory of cluster position. Overlaying this, the momentary production of sounds (in a user-specified window of time) is plotted in red. Therefore, the image in FIGURE 8A (a movie snapshot) provides both the history and momentary state of a cluster. In the case of FIGURE 8A, the momentary state shown is near the end of vocal ontogeny (at 90 days), and so the red dots fall on the clusters that have developed by this time. This type of movie may also be played in reverse, keeping the memory on. In this case, the yellow dots now represent the final target state that clus-

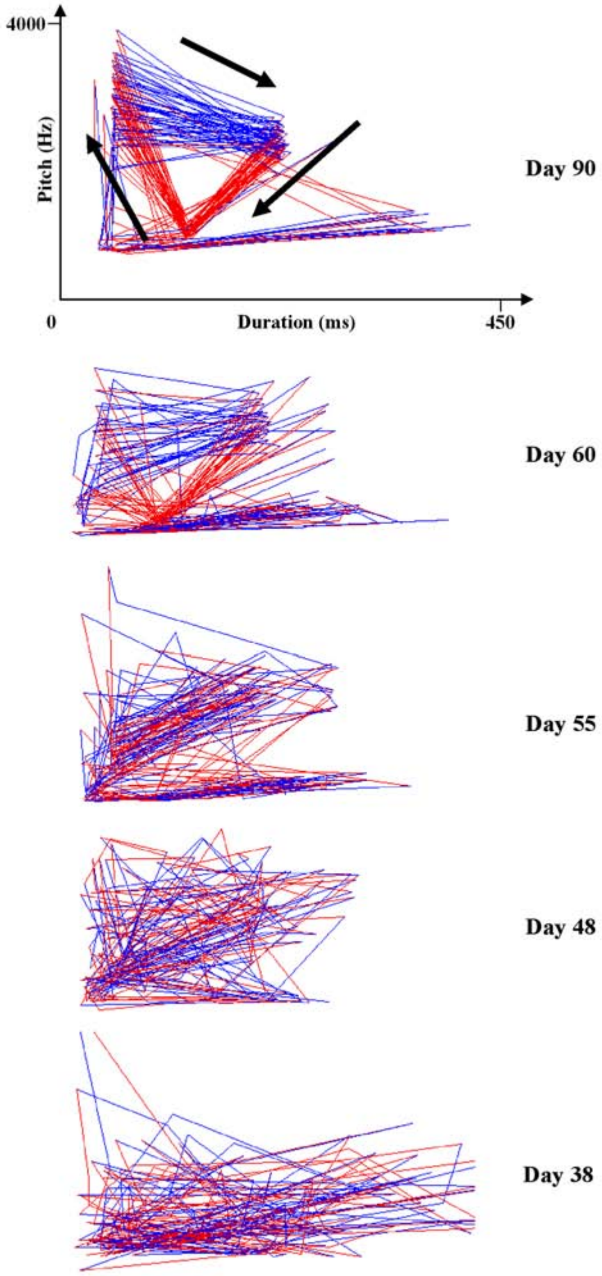


FIGURE 6. Snapshots of DVD-maps showing syntax development in one bird. Instead of plotting the clusters, we plot the trajectories that connect them to illustrate the sequential order of syllables within the song bout. Trajectories to the right (a sequence of a shorter-duration syllable followed by a longer-duration syllable) are denoted by blue lines and trajectories to the left by red lines (color shown online).

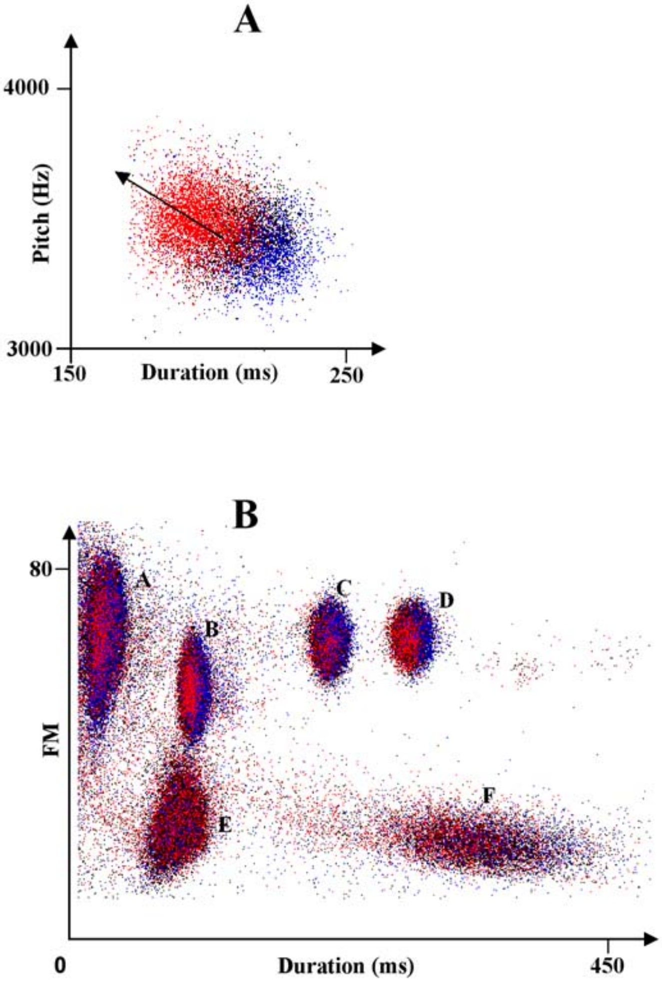


FIGURE 7. Color coding time in DVD-maps [color shown online]. Earlier-produced sounds are represented by blue dots and later-produced sounds by red dots. (**A**) Color representation of developmental time. The *arrow* indicates the direction of the movement of the cluster during one week of song development. (**B**) Color representation of circadian time: syllables denoted by red color occurred during the morning, whereas the blue dots stand for syllables produced during noon time of the same day.

ters will achieve, and the red dots represent sounds produced during an earlier window of time. In FIGURE 8B the yellow cluster on the left represents a sound that has not yet formed. By viewing this DVD-map movie, running the window of momentary state forward or backward in time, it is possible to see when, and how quickly, this time-warping occurred.

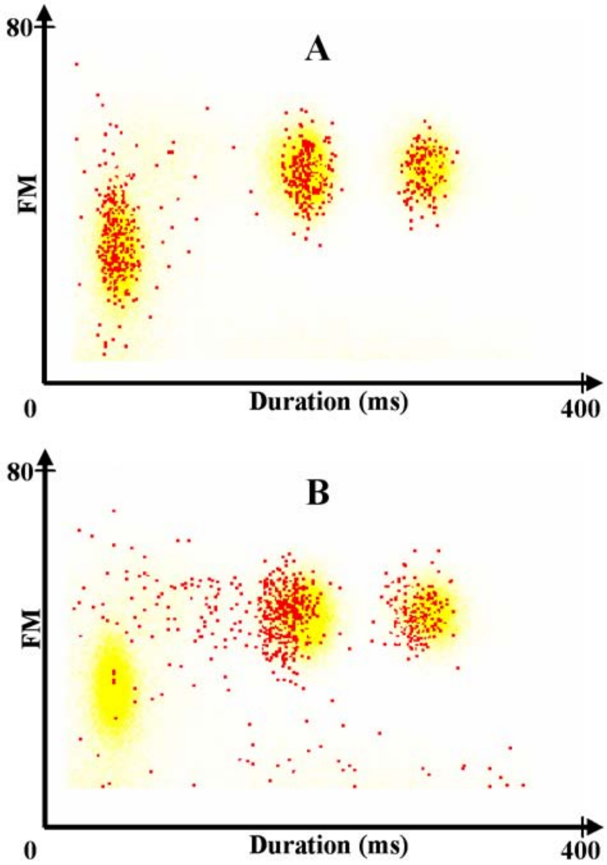


FIGURE 8. Representing different time scales on DVD-maps [color shown online]. (**A**) The feature values of previously produced sounds are plotted in yellow, providing a long-term memory of cluster position; the momentary production of sounds is plotted in red. (**B**) The yellow cluster on the left represents a sound that has not yet formed. The red dots (showing the momentary state at 60 days) indicate that the other two clusters have already been generated by 60 days, although the position of the red dots is displaced to the left, indicating that a time warping occurred for these sounds between day 60 and day 90.

CONCLUSIONS

Understanding the rules that govern development—whether of cells, circuits, or songs—using only static images is like trying to intuit the rules of soccer from examination of hundreds of snapshots taken at different times during different games. Developmental biologists are now using filming techniques to study the dynamic processes underlying such events as axonal path finding and synapse formation. As one investigator put it, “humans can take in lots of visual information at once and extract patterns from it; complex images and movies provide such information.”[12]

We have applied an analogous strategy to the study of song development, creating sets of graphic representations that can be used to present song data from an entire vocal ontogeny. The methods we used are based on simple features of sound and so provide intuitive ways of exploring vocal changes in a large data set. Our system uses the sonogram image, which is much more detailed, to further explore the details of vocal changes, once DVD-maps have captured and characterized them.[14] With the approaches presented here, detecting vocal changes is now much easier and one can follow, more or less in real time, how learning progresses throughout each day. While our approach has focused on song development, its utility for other aspects of birdsong research is readily apparent. For example, most studies on the effects of experimental manipulations, such as lesions, upon stereotyped adult song production have used only sporadic sampling of the effect of an experimental treatment, for example, before and after comparisons, using a limited number of endpoints. The approach outlined in this chapter would allow the experimenter to follow the unfolding of changes in song structures from its initial disruption to its eventual recovery—shedding light on both processes. DVD-maps updated in real time could be particularly useful in the context of neural recording or molecular experiments, where data or tissue collection must be timed to coincide with particular events during the vocal learning process. Extending these tools further, one could attempt to generalize the DVD-maps to integrate, within a single accessible display, correlated data on song structures, neural firing, and behavioral state. The methods described in this chapter should facilitate the extension of such approaches to a wide range of problems in birdsong neurobiology.

ACKNOWLEDGMENT

This work was supported by a Public Health Service Grant DC004722-05 to O.T. and by a National Institutes of Health RCMI grant to CCNY.

REFERENCES

1. Barrington, D. 1773. Experiments and observations on the singing of birds. Phil. Trans. R. Soc. Lond. **63:** 249–291.
2. Thorpe, W.H. 1958. The learning of song patterns by birds, with especial reference to the song of the chaffinch *Fringilla coelebs*. Ibis **100:** 535–570.
3. Catchpole, C.K. & P.J.B. Slater. 1995. Bird Song: Biological Themes and Variations. Cambridge University Press. Cambridge.
4. Hauser, M. & M. Konishi. 1999. The Design of Animal Communication. MIT Press. Cambridge, MA.
5. Ho, C.E., B. Pesaran, M.S. Fee & P.P. Mitra. 1998. Characterization of the structure and variability of zebra finch song elements. Proc. Joint Symp. Neural Computat. **5:** 76–83.
6. Tchernichovski, O., F. Nottebohm, C.E. Ho, *et al.* 2000. A procedure for an automated measurement of song similarity. Anim. Behav. **59:** 1167–1176.
7. Baker, M.C. & D.M. Logue. 2003. Population differentiation in a complex bird sound: a comparison of three bioacoustical analysis procedures. Ethology **109:** 223–242.

8. Cortopassi, K.A. & J.W. Bradbury. 2000. The comparison of harmonically rich sounds using spectrographic cross-correlation and principal coordinates analysis. Bioacoustics **11:** 89–127.
9. Janata, P. 2001. Quantitative assessment of vocal development in the zebra finch using self-organizing neural networks. J. Acoust. Soc. Am. **110:** 2593–2603.
10. Derégnaucourt, S., J.C. Guyomarc'h & V. Richard. 2001. Classification of hybrid crows in quail using artificial neural networks. Behav. Process **56:** 103–112.
11. Tchernichovski, O., T. Lints, P.P Mitra & F. Nottebohm. 1999. Vocal imitation in zebra finches is inversely related to model abundance. Proc. Natl. Acad. Sci. USA **96:** 12901–12904.
12. Beckman, M. 2003. Play-by-play imaging rewrites cells' rules. Science **300:** 76–77.