Mechanisms of Vocal Learning in Songbirds:
how song syllable structure is learned

by

Primoz Ravbar

A dissertation submitted to the Graduate Faculty in Biology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2012

This manuscript has been read and accepted for the Graduate Faculty in Biology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Dr. Ofer Tchernichovski________________

____________________________ ____________________________

Date Chair of Examining Committee

Dr. Laurel Eckhardt _____________________

____________________________ ____________________________

Date Executive Officer

Dr. Josh Wallman ________________________________

Dr. Lucas C. Parra________________________________

Dr. David S. Vicario________________________________

Dr. Michael A. Long________________________________

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

by

Primoz Ravbar

Advisor: Professor Ofer Tchernichovski

Co-advisors: Professor Lucas C. Parra and Professor David S. Vicario

When learning to perform continuous actions one needs to cope with conflicting motor requirements: while some parts of the action may require exploratory variability to find motor states that can efficiently produce a desired outcome, other parts may require consolidation if they are already close to the desired goal. A possible solution to this problem is partitioning the action into segments that could be controlled independently, but it is not known if exploratory variability can be locally regulated during the sensory-motor learning of a continuous action. This dissertation explores song learning in songbirds as a model of learning to perform continuous actions. In the first set of experiments we manipulated song learning in zebra finches (*Taeniopygia guttata*) to experimentally control different requirements for vocal exploration in different parts of the song. We first trained birds to perform a one-syllable song, and once they mastered it we added a new syllable to the song model. Remarkably, when practicing the modified song, birds rapidly alternated between high and low acoustic variability to confine vocal exploration to the newly added syllable. Analysis of exploratory variability within syllables revealed that acoustic variability changed independently across song elements that were only milliseconds apart. The variability of each song element decreased as it approached the target, correlating with momentary local error and less so with the global error (the mean of local errors across the syllable). We concluded that vocal error is computed locally in sub-syllabic time scales and that song elements can be learned and consolidated independently. However, our evidence for partitions singing behavior into segments that may be learned independently, also indicate that those segments are not developmentally stable. For example, we observed that the appearance of distinct syllable types usually precedes the appearance of distinct sub-syllabic song elements, suggesting that the time scales at which syllables are learned might get shorter over development, perhaps in a hierarchical manner, from the learning of course temporal structure to the learning of fine temporal structure. To test this hypothesis we studied song learning at the articulatory level by measuring the control of respiratory pressure during song learning. We found that during early song learning, respiratory pressure patterns were coarse, and fine temporal structure was later added to that pre-existent coarse structure. This effect was not observed in socially isolated birds that developed their song without learning a tutored model. We therefore propose that the learning of continuous singing action is achieved by first partitioning the song into coarse segments, and then to finer units, where exploratory variability in each unit is dynamically gated until a local match to the song model is achieved.

Acknowledgments

Many thanks to my advisor Dr. Ofer Tchernichovski and Co-advisors Dr. Lucas C. Parra and Dr. David Vicario for their excellent supervision of my work. Thanks for being patient as well as impatient with my progress. I also want to thank Dr. Josh Wallman and Dr. Michael A. Long for their service as Examination Committee members and for reading of this dissertation.

Big thanks to Dina Lipkind, Franz Goller (and Goller lab members), Brent Cooper, Pratik Shukla and Xiao Jianqiang for their critical contributions to data collection and experimental design. Julia Hyland Bruno helped me in the darkest moments of my English breakdowns with proofreading. Thanks! Tina Roeske helped a great deal with quality consultation. She is a passionate debater, a devil's advocate and a motivation booster at the same time. Thanks to David Vicario with whom I had some of the most profound and engaging conversations about everything under the sun and beyond, while driving back and forth between Rutgers University and New York City.

I would like to thank all present and former members of Vicario and Tchernichovski labs for sharing their ideas, successes and frustrations with me.

Thanks to everyone in the Biology Department of GC CUNY for extensive administrative help with this dissertation.

Thanks to my parents for letting me crossing the streams where there were no bridges.

Thanks to Rui Hosono for encouragements and spiritual support.

Dedicated to Josh Wallman whose mastership of thought and life inspires my infinitely.

Table of Contents

Chapter 7: Integration and Discussion…………………………………………………….102

List of figures

Chapter 1: Introduction

This dissertation studies the learning of continuous actions. I will start with a brief review of discrete learning of goal-oriented actions, using an example of throwing darts at a target to present the question of the trade-off between exploratory variability and consolidation, which arises from simple reinforcement learning of goal-oriented actions. This will be followed by a presentation of the challenges associated with learning of more complex continuous actions such as dancing. We will see how the time scales of motor exploration and error evaluation are important in the learning of such actions, and examine how exploratory variability can be applied to the learning of continuous actions, where some parts of the action may be already learned while other parts may require more exploration. I will discuss possible solutions to this problem, and focus on the partitioning of the action into segments that could be evaluated separately (local error assessment) as a potentially effective mean of simplifying the learning task. I will propose testing this hypothesis in songbirds, where song learning as already been established as a suitable model system for testing such hypotheses, which will bring us to the main aims of this dissertation:

1. Test if birds learn to imitate song syllables (continuous action) by computationally partitioning them into segments. This way the error can then be assessed locally in each segment and exploratory variability can be confined to those song elements that need to change most, while other song elements (those that are already well imitated) can consolidate. Testing this hypothesis required developing of experimental and statistical methods to measure exploratory variability of individual song elements in time scales of

milliseconds and across hundreds of thousands of syllables produced during development.

2. Test if the units of segmentation might change during development. If so, what are the implications?

1.1 Reinforcement learning and the role of exploratory variability

Reinforcement learning can be described as a simple form of learning where an animal associates an action with a stimulus that follows the action. If as a consequence of presenting the stimulus the frequency of repeating that action increases then we can conclude that the stimulus is internally evaluated as a reward (positive reinforce)(Watson, 1913, Skinner, 1938; Ferster and Skinner, 1957). The concept of reinforcement learning was further developed in other areas such as machine learning, dynamic programming and control theory (Kaelbling et al., 1996). The "goal" of the animal in reinforcement learning is to maximize the cumulative reward (the positive reinforcement stimulus), which could be presented to the animal either internally or externally. In the simplest case of reinforcement learning the animal starts with no prior knowledge about how actions are rewarded. Then, by trial and error the animal accumulates knowledge about which actions resulted in highest rewards, and consequently the rate of well rewarded actions increases.

Reinforcement learning requires both motor exploration (trials) and consolidation of learning (reducing the error), and there is often tension between these two requirements as can be illustrated by the following three cases of simple task learning: Imagine you are looking for a golden wrist watch buried somewhere in send on a beach, using a metal detector. The instrument will emit pulses of sound at increasing rate (reward) as you approach the watch (the target). In this trial and error task you move the detector, either systematically or randomly, to different locations while listening for the rate of sound pulses. Once pulse rate goes above baseline you try to lock on the target: If the rate is higher than in a previous location, then you know that the move has been in the right direction. As you get closer to the target you would typically reduce the amplitude of each move in order to prevent overshooting the target. Therefore you decrease the magnitude of exploratory movement as the rewards get higher (the rate of sound pulses from the metal detector increases). Conversely, if the reward is low there is little risk of reducing it even further by changing the position of the metal detector and consequently the magnitude of exploratory movements is usually higher.

Now consider a somewhat different example: throwing of darts to a target. The goal is to hit as close as possible to the center. In this case you know the location of the target but it requires considerable practice to achieve good performance. Computing the coarse trajectory of your movement is done instantly (Kawato, 1999), but there is a prolonged process during which you are exploring different variations of throwing movements. You are then “selecting” those trajectories of throwing movements that improved your performance; you try to repeat the throwing movements where you know you were “doing something right” and try to improve them even further. This way the variability of movements should decrease as darts land progressively closer to the target. You can evaluate this distance and estimate the error with each throw.

Finally, consider a case that combines both cases of simple task learning described above. Imagine you are throwing darts into a white screen behind which the target is hidden. Suppose there is a metal detector at the location of the invisible target and it responds with increasing rate of sound pulses as the (metal) darts land closer to it, i.e. *error signal* decreases. In this case you are exploring the surface of the screen as well as the parameter space of movement trajectories that will result in more accurate and more precise throws. As before, the goal of learning of this task is to maximize the cumulative reward (the rate of increase in sound pulses across throws). Initially the trajectories of the throwing movements are variable but over the course of learning those trajectories that resulted in a decreased error signal are more likely to be repeated and variability consequently decreases. Once the error signal stops decreasing (the cumulative reward of several throws is constant) you effectively turn off the *exploratory variability* and your throwing movements become stereotyped (consolidated).

The cases above all describe *goal-oriented* actions, where a particular trajectory of the throwing movement taken does not matter, as long as it results in a decreased error signal. The trade-off between exploratory variability and consolidation in goal-oriented actions can be therefore formulated as following: when the error signal is high (rewards are low) it pays to explore different throwing movements that could potentially result in higher reward, however, if the cumulative reward of several throws is relatively high than exploration becomes more risky (as it could result in decreased reward when the darts land further from the target). Consequently the gain of exploratory variability decreases (Kaelbling et al., 1996; Sternad and Muler, 2009).

1.2 The role of exploratory variability in learning of continuous actions

Most sensory-motor learning studies are concerned with learning of discrete goal-directed movements, such as reaching (van Beers, 2009) and throwing movements, as exemplified above (Muler and Sternad, 2009). In goal-directed actions the success of learning is often estimated by a single parameter (a vector, or a scalar value), such as the distance of darts from the target. Therefore the error estimate is *global*, i.e., the execution of the *entire action* is evaluated against a single error estimate, which includes two parameters: angle and distance from the target. The error in turn is estimated only at the last time-point of the action, called the "end-point". In the case of throwing darts this end-point would be the vector describing the velocity and the angle of a dart at the moment when it is released from the throwers hand (in the absence of environmental noise, this vector will fully determine where the dart will land with respect to the target). Of course, with four joints involved, there are many possible kinematic trajectories that might arrive to the same end-point vector (and thus the same error estimate). Therefore, it is not trivial even in simple task that a global error estimate is sufficient to allow efficient learning of goal-directed movements. Nevertheless, there is evidence that error could be estimated continuously. Many of the studies that show this utilize the force field experiments in which perturbations are applied to the subject's hand during a reaching movement (Shadmehr and Mussa-Ivaldi, 1994). Reaching movements (and other goal-directed movements such as throwing of darts) are under ballistic control, meaning that they are computed (internal model of limb dynamics is created) before they are executed and that sensory feedback has no immediate (on-line) affect on the trajectory of their execution (Morasso, 1981). Experimentally induced perturbations are normally a function of hand position and/or velocity (or acceleration). Before the perturbations are applied the movement trajectory is typically a straight line (going right for the goal). As soon as the perturbations are presented to the force field the movement trajectory deviates from the straight line. But with practice the trajectories start converging back to the original straight line (compensation). This learning is possible because the internal model of the force field and limb dynamics can change adaptively, e.g. it can predict the change in the force field after the perturbation has been applied (Bhushan and Shadmehr, 1999; Kawato, 1999).

What is the nature of the error signal? In two cases presented above the error signal was non-parametric – the learner only know how far was the target but not its direction. Even in the case of throwing darts to a visible target we can imagine a naïve learner how has no model of limb dynamics at all and has to randomly explore trajectories that will bring the darts closer to the target. But such non-parametric error signal might not be realistic in animals. We do not randomly vary the trajectories of throws when trying to hit the center of the target. Rather, the exploratory variability employed has a direction such that the throws vary more in the direction of the visible target. This is because the error signal is informative and the learning process guided (Andalman and Fee, 2009; Engel and Soechting, 2011).

Learning of discrete (“simple”) goal-oriented actions could be contrasted with learning to perform *continuous actions* such as driving a car or performing a dance. In continuous actions the goal of learning is not the set of parameter values at the end-point, but rather the entire trajectory of the action. Thus the quality of a dance performance, for example, is not evaluated at any particular moment in time but along the entire performance. So how is the error signal produce when learning continuous actions and how is the exploratory variability applied? There are a few possible scenarios illustrated by the examples bellow.

Imagine a dance student who is learning to perfect her act. The dance teacher evaluates her performance at the end of each act, but only with a numerical grade (such as 1-10). The teacher never explicitly points out any particular weakness of the act and the student randomly varies her performance across all elements of the act. Now in real life the evaluation of performance would rarely be non-parametric. The error signal would have a direction. For example, the student could watch a video of the performance she wishes to imitate and compare it to a video of her own performance. The error estimates would then have a direction. But as will be discussed later the nature of the error signal in learning of birdsong, as an example of continuous action, is still somewhat an open question. In order to present the main problem studied in this dissertation we will employ a simple (even if unrealistic) case of a non-parametric error signal such as grades from 1 to 10.

So in this case the student selects those acts that have received higher grades and is more likely to repeat them (as in the case reinforcement learning of throwing of darts), while still retaining some exploratory variability, hoping to improve the current grade. In this case of continuous action learning the error is evaluated *globally* (the sum of errors across all time-points of the dancer's trajectory) and the exploratory variability is applied across the entire act because the student does not know which particular part(s) of the act should be improved and which ones should change less. Consequently, the student will at times change the part of the act that had been already perfected and thus somewhat deteriorate the performance. This case therefore illustrates the conflict between exploratory variability and consolidation that can arise from learning of continues actions. Whereas some parts of the action may require exploratory variability to find motor states that can efficiently produce a desired outcome, other parts might require consolidation if they are already close to the desired goal. Even so, learning of continuous actions with only global error estimates can work, as has been suggested by models of song-learning in birds (Fiete, 2007), where combining uniformly distributed exploratory variability with a mechanism for comparing the overall (global) similarity to the song model that a bird is attempting to imitate is theoretically sufficient to enable song learning. It has also been experimentally confirmed that birds can improve individual song elements even if the error estimate is global, as in our example of the dance student (Charlesworth et al., 2011). In these experiments a negative reinforcement was applied when fundamental frequency of a certain element of the song did not reach a threshold specified by the experimentalists. Importantly, the negative reinforcement could be applied at any time during or after the song performance. This result will be discussed further in Chapter 7.

But how can be such conflicting demands between exploration and consolidation satisfied during learning? Imagine again the dance student who now finds a new teacher. This time the teacher explicitly grades separate elements of the dancing act (again with a numerical grade). Thus the parts of the act that the student needs to improve will receive a lower grade (high error signal). We can say that the error is now evaluated *locally*. With this information at hand, the student can apply exploration to different elements of her act separately, so that those elements that have received high grades become consolidated while elements of the act that need improvement can vary more. Now her dancing act does not have to deteriorate when more exploration is applied.

So rather than evaluation the performance of a continuous action globally, an alternative approach would be to *partition* the task into several short segments, compute local errors, and approximate the target in a piecemeal manner. The segmentation of continuous action during reinforcement learning has been studied by several theoretical models (Doya, 2000). It has been shown by the models that either in the case of global or local error estimates, the gain of exploratory variability must decrease with learning, but partitioning the task to discrete segments could also make it useful to relive the tension between the conflicting requirements for exploration and consolidation in different parts of the action. However, while it is feasible that animals do employ partitioning of continuous actions, there is no direct experimental evidence to it. In this dissertation we use birdsong as a model for continuous action learning and show that, indeed, zebra finches can locally evaluate the elements of their song during the course of learning and can apply exploratory variability to those parts that need to change most. In the following section I will present my considerations in using birdsong as a model for continuous action learning.

1.3 Zebra finch song as a model for learning of continuous actions

The song of adult zebra finches is composed of bouts of repeated units commonly called “syllables” (see Figure 1.1, day 90).

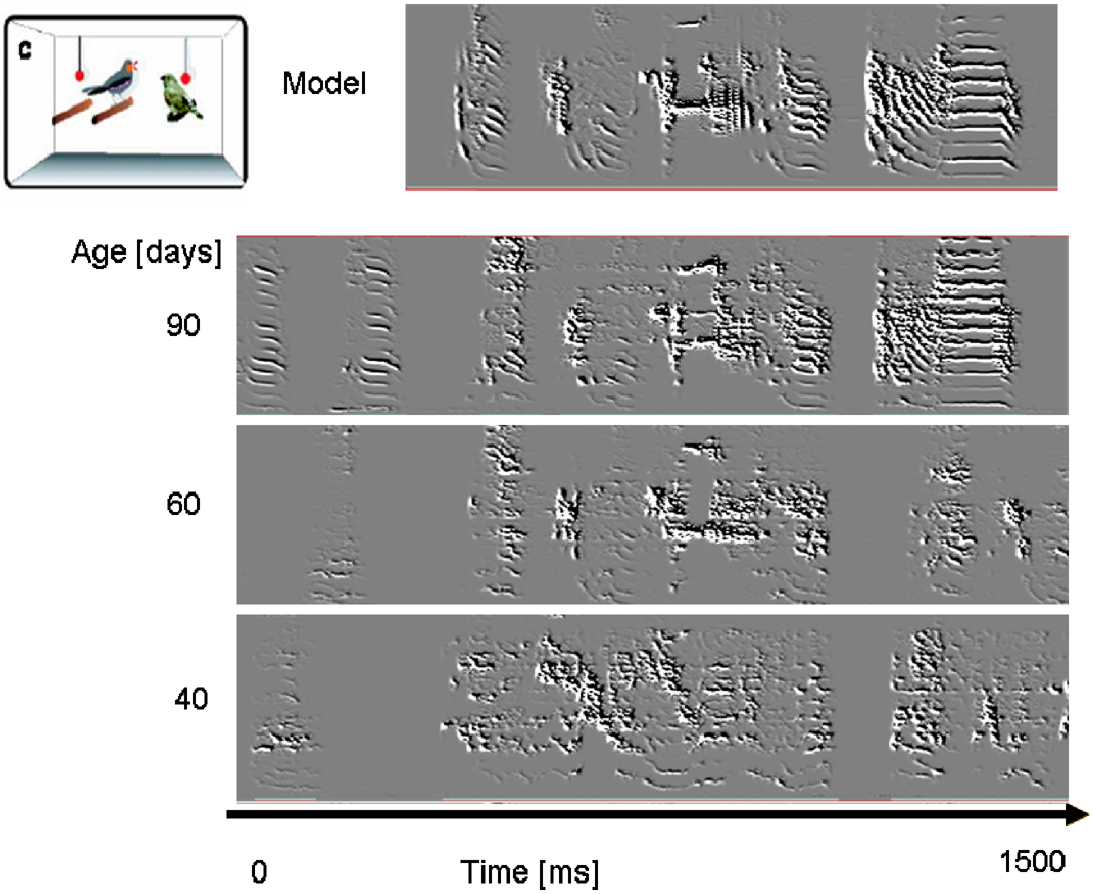

Figure 1.1 Structure and development of zebra finch song. Birds in acoustically isolated boxes can be tutored by a model song played from a speaker. The sonograms show how the song develops to resemble the model (from 40 to 60 days of age). The fully "crystallized" song bout, shown at day 90, has three syllables and two "introductory notes" before them.

Typically a song bout consists of 2-5 different types of syllables proceeded by a few shorter "introductory notes". Zebra finches are frequent singers and can produce up to 30,000 song syllables per day. The order of syllables in a bout is quite conserved in a fully developed song, although skipping of a syllable is not uncommon. If interrupted, a bird will stop singing at the end of a syllable rather than break the song within a syllable (Franz and Goller, 2002; Cynx, 1990). In this sense syllables can be understood as discrete units rather than a fully continuous action. A typical duration of a syllable is about 130-280 ms and it consists of 4-6 vocal elements commonly called “notes”. During song development it is technically easier to detect distinct time events within syllables rather than segmenting the syllables, and therefore, we will refer to intra-syllabic structures (notes and events) as “vocal elements”. Unlike syllables in a song bout, the vocal elements within a syllable are invariably performed in the same order and there are rarely distinct boundaries between them, and as such a syllable could be considered a true continuous action. This dissertation will mostly focus on the development of syllabic types than include several vocal elements.

The ability to train zebra finch to perform specific syllables, and to record their entire learning (every single performance of each sound) make them a nearly ideal model for studying sensory-motor learning of continuous action and for studying the role of exploratory variability in that learning; We were able to follow developmental trajectories of different vocal elements separately and reliably identified these vocal elements during prolonged developmental epochs. The shear amount of singing with thousands of repetitions per day promises robust statistical analysis.

Neurobiology of song learning mechanisms

Song production and learning in male zebra finches involves in two brain pathways: anterior frontal pathway (AFP) involved in learning (Figure 1.2, red color) and the production pathway, which includes nuclei HVC and RA (Figure 1.2, blue color). Only the main nuclei of these pathways are illustrated in Fig. 1.2. Lesions of AFP prevent song learning (Bottjer et al.,1984; Scharff and Nottebohm, 1991; Brainard and Doupe, 2000; Haesler et al., 2007) but have less affect on song production, while lesions of RA or HVC (in adults) completely prevent song production.

Although birds can sing without AFP their song can no longer change after the lesioning (Brainard et al, 2000, Haesler et al, 2007) and its performance becomes extremely stereotyped (Scharff and Nottebohm, 1991; Olveczky et al., 2005). As we will see later, even in the song of normal adult zebra finches there is still some residual variability present, but in birds with lesioned AFP even this variability is minimal. This finding suggested that the variability might be functionally connected to learning. Then came the direct evidence that variability of song patterns can be used for vocal exploration (Olveczky et al, 2005, Andalman and Fee, 2009). In these studies AFP was temporally inactivated by injections of TTX toxin to the AFP nucleus LMAN (Figure 1.2). The injections were done during the sensitive period of song learning and resulted in the complete ablation of variability while the structure of the syllables seemed to revert to the developmental stage of the previous day (Andalman and Fee, 2009). This result is important for understanding of the nature of the error signal. If the error signal was entirely non-instructive, resulting in exploratory variability with no particular direction, then the structure of the song would not revert to previous developmental stage during LMAN inactivation. The song would become stereotyped but its structure would not change. This suggests that AFP is providing some information about the direction in which the structure of the song should change.

In young birds the singing of very variable and unstructured subsong is driven by AFP pathway as the bilateral lesions of HVC do not prevent singing (Aronov, 2008). While these variable song patterns dominate singing behavior in juvenile birds, during development neural control gradually shifts to a second vocal center called nucleus HVC (a proper name) (Aronov, 2008). In contrast to AFP, the neurons in HVC generate highly stereotyped electrophysiological activity (Hahnloser et al, 2002; Kozhevnikov and Fee, 2007). And, as noted earlier, in the absence of AFP pathway the activity of HVC pathway results in a very stereotyped song production. Both variable song patterns from AFP and stereotyped song patterns from HVC converge in the premotor song nucleus RA which, in turn, controls primary motor nuclei in the brainstem that drive the muscle systems involved in song production (respiratory muscles and the muscles of the syrinx) (Schmidt, 2004). Consequently, as HVC gradually takes over the control of song production during development, the acoustic variability of the song decreases until the song becomes fully crystallized with only a small gain of residual variability originating from AFP. This residual variability can still be used to modify the song even in older birds (Tumer and Brainard, 2007).

.

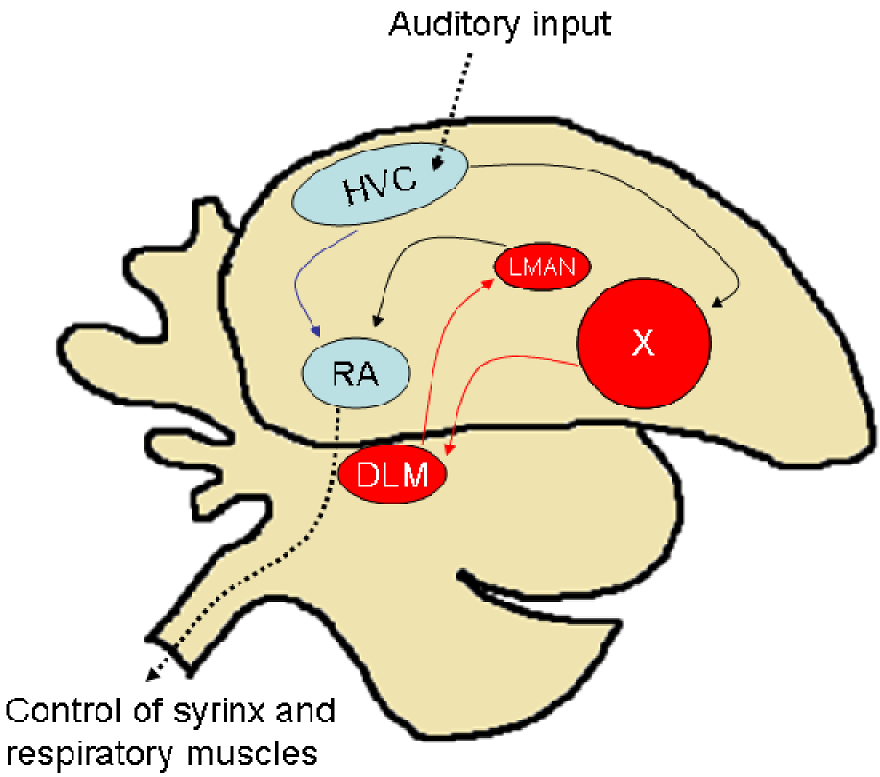

Figure 1.2 The song system. Two pathways are responsible for song production (blue nuclei) and learning (red nuclei). The HVC nucleus (proper name) is involved in both, sensory and motor processing. The HVC → RA connection is necessary for song production as lesions of either HVC or RA prevent the ability to sing in adult animals. The red nuclei represent the Anterior Frontal Pathway (AFP) which is necessary for song learning. LMAN nucleus from AFP projects onto RA. Lesions in LMAN prevent song learning and result in very stereotyped song production.

Song variability can also change adaptively in short time scales, depending on social context and behavioral state (Brainard and Doupe, 2000; Sakata and Brainard, 2009; Jarvis, 1998; Kao, 2005). During courtship, males sing to attract females and such female-directed song is significantly less variable than the undirected song. This effect is particularly strong in juvenile birds (Hessler and Doupe 1999 a,b). A possible interpretation of this result could be that birds do not explore the acoustic space (they are not engaged in active learning) while they perform to females. Exploratory variability present in the female-directed song might negatively affect its structure. In contrast, during undirected singing (practice) the song is variable as the bird is engaged in vocal exploration. It has been shown that practice singing triggers a strong activation of immediate early genes at the AFP (Jarvis, 1998) and consequently the premotor activity in AFP (nucleus LMAN, projecting to RA) becomes more noisy and the song more variable (Kao, 2005). During female directed singing, on the other hand, there is no apparent activation of early gene expression in AFP and the premotor activity becomes synchronized. Although at the macro level AFP output is noisy during practice singing, microstimulations of LMAN neurons result in brief and very specific, time dependent modulation of song features (Kao, 2005). Therefore AFP can play a role in reinforcement learning by “injecting” exploratory variability to the RA at narrow time-scales.

Although this dissertation does not include an investigation of neuronal mechanism, the behavioral results will allow us to present alternative hypotheses about the possible role of AFP in regulating vocal exploration.

1.4 Hierarchical learning of complex actions

The central question of this dissertation is if vocal exploration might be locally regulated by computing deviations (errors) from the song model in short segments of the song, namely, weather the exploratory variability could be regulated at short time-scales. If indeed the bird computationally partitions the song to several segments, how should the sizes of segments be determined?

To illustrate this problem, let us go back to the case of a dance student. Imagine that her new teacher evaluates the dancing act at just a few time-points, say three times during the whole act (broad segmentation of action). The student only needs to remember the three grades that she receives after every act and then apply exploratory variability accordingly. But because the partition of the dance act is so broad, within each segment the same conflict between consolidation and exploration arises as with global error estimates. On the other hand, the teacher could evaluate the dancing act at very short intervals (say 100 times per performance). This would create a new problem: difficulty to remember the whole grade vector (100 grades) and track each segment separately. It would also mean more work for the teacher, so when we are discussing reinforcement learning with *internal* reinforcers, as could be the case in birdsong (Fiete, 2007) the error evaluation would present an additional challenge.

One solution to the problem of segment size could be *hierarchical learning*. Initially, when error is large, the risk of exploration is small (remember the trade-off between exploration and consolidation presented above). Therefore, in the beginning of learning period the segmentation can be broad. Later, as the overall performance improves, the segmentation can be narrower.

In our case of the dance student, imagine her initial performance is very far from where it should be. Most elements of her act are far from the target. It would not make much sense for the teacher to grade (again numerically) each element. Instead a global low grade would be given.

Now, as the student varies her act, suppose the first half of it becomes better (by chance). It would become worthwhile to decrease the amplitude of exploration in this part, while continuing changing the second part. So two grades could be given (a higher one for the first part) and the act would be partitioned into two. Imagine repeating this process of partitioning the act to shorter and shorter elements. We will refer to this gradual decrement of segment size as "structural refinement".

We have observed such increasingly fine partitioning of the song when we analyzed not the acoustic structure of the song, but rather the structure of respiratory pressure. This pressure, produced by bird's air sacs drives the singing as it forces the air through the syrinx (Goller, 2002; Suthers and Margoliash, 2002). But unlike in human speech, the structure of the respiratory pressure in a singing bird can be relatively modulated, e.g. it is made of many short elements. How does such complex structure of respiratory pressure develop?

The results presented in this dissertation (Chapter 5) suggest that initially a bird learns very broad modulation of respiratory structure (coarse structure). But with practice he adds increasingly short finer structure to the pre-existing coarse structure. This suggests that the development of the complex structure of respiratory pressure might indeed be hierarchical as the granularity becomes progressively finer (the action is partitioned into progressively shorter segments).

Chapter 2: Vocal exploration is locally regulated during song learning

Data collection and experimental design used in this chapter was contributed by Dina Lipkind.

2. 1 Background and Rationale

Exploratory variability is essential for sensory-motor learning. We know that variability across renditions of syllables decreases over song development as the structure of the song becomes increasingly similar to the model (e.g. the acoustic error decreases). It has been shown that this exploratory variability is necessary for song learning (Brainard and Doupe, 2000; Olveczky et al., 2005). But it is not known how and at what time scales can variability be regulated. It could be, for example, that acoustic error is estimated globally and as it decreases, the exploratory variability of the entire song decreases with it (global consolidation). On the other hand, observing local consolidation (e.g. where variability decreases locally in parts of the song where the target is approached) would imply that acoustic error is estimated locally.

In order to disambiguate between these possibilities we manipulated song learning in zebra finches to experimentally control the requirements for vocal exploration in different parts of their song. We first trained birds to perform a one-syllable song, and once they mastered it we added a new syllable to the song model. If the birds can regulate the exploratory variability only globally, we predicted that the addition of a new (and unstructured syllable) to the song would result in the increment of variability gain across the entire song bout, including the already mastered syllable. On the other hand, if variability can be regulated locally we predicted that exploration would be confined only to the newly added syllable, which would also imply that acoustic error can be estimated locally.

2.2 Methods

2.2.1 Animal care

All experiments were conducted in agreement with USNIH guidelines and were reviewed and approved by the Institutional Animal Care and Use Committee of City College of New York, City University of New York.

2.2.2 Training procedure

Birds were bred in family cages. Fathers were removed when clutch mates were 7-8 days old or less, and thereafter birds were raised by their mothers and were not exposed to songs. On day 30-32 post-hatch, male birds were individually isolated in sound-attenuation chambers. All birds were tutored with operant song playbacks from day 43 to 90 days post hatch, as described in Tchernichovski et al., 1999. Eight birds used in the study presented in this chapter were trained using an altered-target training procedure as described below.

2.2.3 Song recording and analysis

We audio recorded (16 bit, 44.1kHz) each bird continuously from day 32 to day 90 post-hatch using Sound Analysis Pro 1.4 (Tchernichovski et al., 2000). Recording epochs containing songs were automatically identified and saved, and song features (amplitude, pitch, Wiener entropy, etc.) were computed as in Tchernichovski and Mitra (2004). Multitaper spectral analysis (Tchernichovski et al., 2000) was performed with time windows of 10 ms, advancing in steps of 1 ms such that song features were computed for every millisecond. Syllable boundaries were identified using a stationary threshold of sound amplitude (segmentation).

2.2.4 Altered-target training procedure

Zebra finches were trained sequentially with two song models (“source” and “target”) as in Lipkind and Tchernichovski (2011, Fig 2.2 *A*, *B*). Source and target song models for training were composed from natural syllables. Twenty-eight birds were trained with playbacks of the source song, composed of a single syllable (AAA…) from day 43 post-hatch, and songs were analyzed daily to determine if the source model was imitated. For birds that learned the source before day 63 (n=15), we switched their training to playbacks of the target song (ABAB…). Eight birds that succeeded to learn the target song were selected for analysis. In six of these birds, the novel syllable B was a harmonic stack (see Fig. 2.2 *A*) and in the remaining two birds, a broadband, highly modulated syllable.

2.2.5 Cluster analysis

After the segmentation of song bouts mean acoustic features of the segments were computed. In the acoustic space these segments tend to fall into clusters as shown in the Figure 2.1.

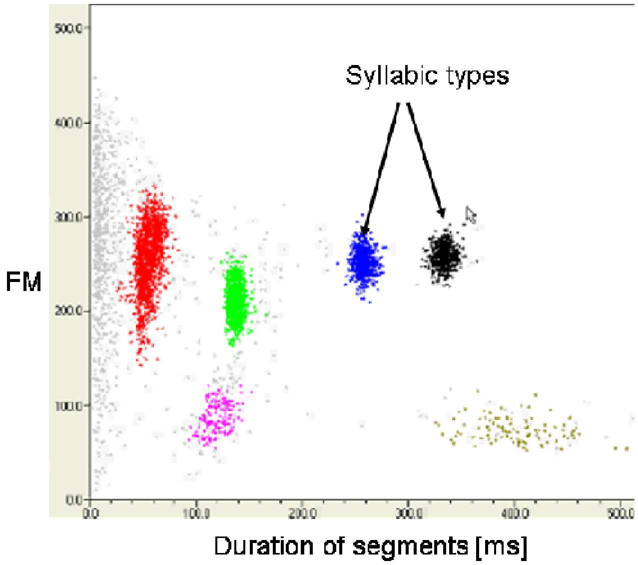

Figure 2.1 Hierarchical nearest neighbor cluster analysis methods was used to identify syllabic types. Song segments typically form clusters in the acoustic feature space. Only two features (FM and duration) are shown in the scatter plot but the clustering algorithm operates on multiple features. Each point represents mean song feature values of a segment. The colored points belong to identified clusters; gray points are not clustered (residuals).

We used a hierarchical nearest neighbor clustering algorithm, using Sound Analysis Pro 1.4 (Tchernichovski et al., 2000) for the cluster analysis. In the figure above identified clusters are colored. Although only two song features are shown (FM and duration) cluster analysis is performed across multiple features.

2.3 Results

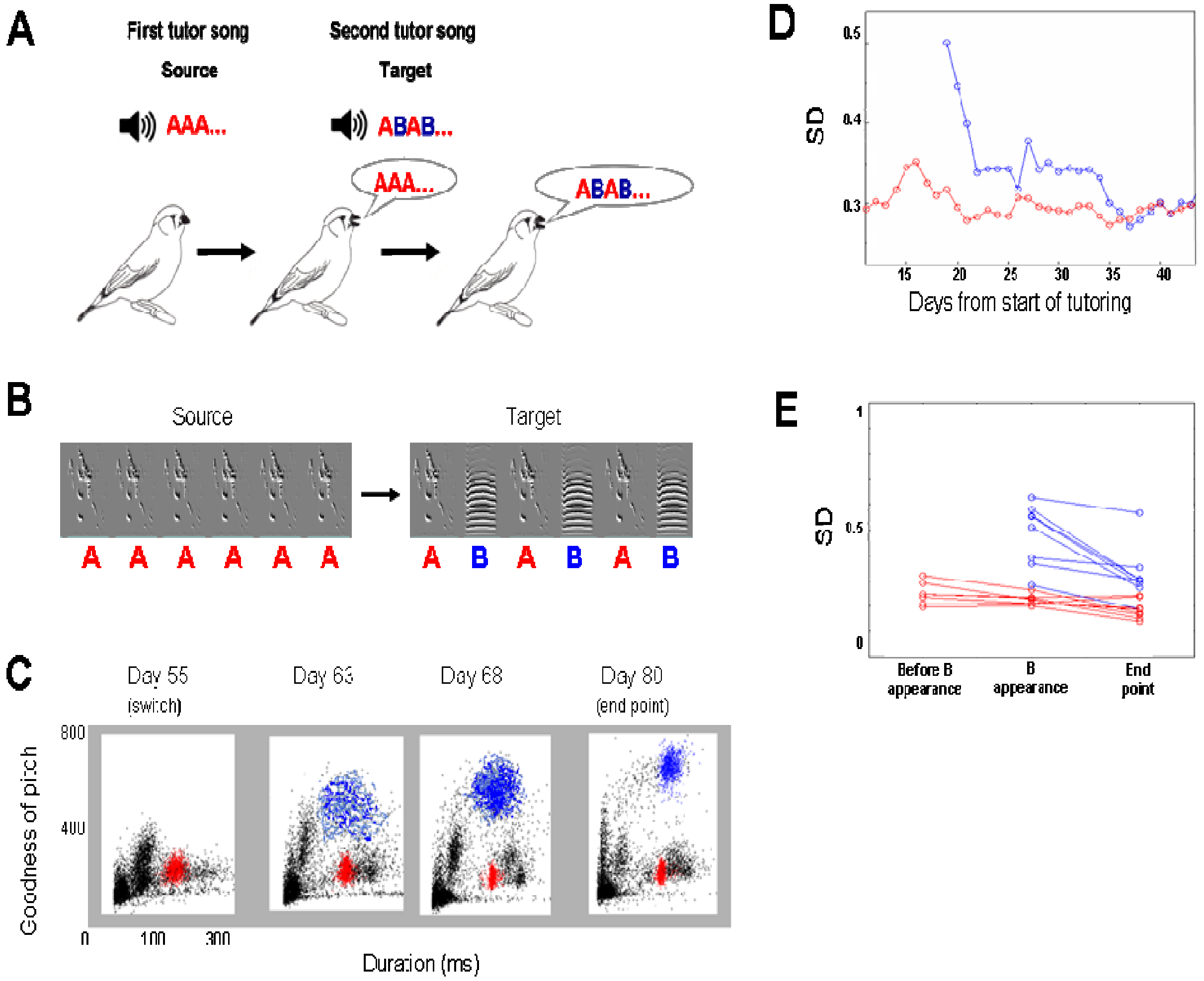

Figure 2.2 Vocal exploration is confined to newly added syllables. ***A,*** The AAAA→ABAB altered target training procedure. ***B,*** Spectral derivatives (sonograms) showing the source and target song models. ***C,*** Scatter plots of syllable features (goodness of pitch versus duration). The red cluster corresponds to syllable A and the blue cluster to syllable B. ***D,*** Variability (SD_{syll}) of syllables A and B across development. This example from one bird shows variability pooled over syllable features. ***E,*** Same as in ***D***, but across birds (n=8), 3 days prior, 3 days after the appearance of cluster B, and last 3 days before the end-point. Note, the variability of A cluster does not change after the appearance of B ($p>0.98$), while variability of B drops significantly ($p<0.003$, single-tailed *t-test*).

2.3.1 Inclusion of a new syllable to the song bout

We first manipulated song learning so that only one part of the song would require vocal exploration. We used an altered-target training approach (see 2.2.4 and Lipkind and Tchernichovski, 2011), training juvenile zebra finches (day 43 post-hatch) first with a source song model (AAAA) consisting of a single syllable and then altering the training to a target song model (ABAB), which included an additional new syllable B (Fig. 2.2 *A*,*B* and Methods). Most birds succeeded in inserting the new syllable into their song bouts and started to produce the target song.

2.3.2 Recording and analysis of the syllables

We recorded and analyzed the entire vocal output of each bird during the transition (AAAA→ABAB), automatically segmenting the songs into syllables. The structure of each syllable was summarized by four features: duration, mean Wiener entropy, mean frequency and goodness of pitch (Tchernichovski et al., 2000). We then performed cluster analysis of syllable features in order to identify the A and B syllable types (see Methods).

Figure 2.2*C* presents an example of one bird, showing scatter plots of two syllable features (duration versus mean goodness of pitch) in different stages of song learning. By the time we altered the tutoring, the cluster that corresponded to syllable A (red) was already small and dense. In contrast, the cluster that corresponded to the newly learned syllable B (blue) was initially much larger and highly scattered. The variability across renditions (SD_{syll}) of syllable B (blue cluster in Fig. 2.2*C*) then gradually decreased until it became similar to the variability of syllable A (Fig 2.2*D*). Interestingly, there was no apparent increase in the variability of syllable A when syllable B appeared (Fig 2.2*C*), indicating that when the bird sang ABAB it rapidly altered between performing a highly stereotyped and a highly variable syllable.

2.3.3 Exploratory variability is confined to the newly included syllable across birds

To test across birds (n=8) if variability of syllable A was affected by the appearance of new, highly variable, syllable B, we calculated the variability (SD_{syll}) of both clusters during three time periods: just prior to the appearance of B (days -3 to -1), just following the appearance of B (days 1-3) and at the end-point (days 90-93 post-hatch). As shown in Fig. 2.2 *E*, the variability across renditions (SD_{syll}) of syllable A (red traces) did not increase during the three-day period after the highly variable cluster B (blue traces) emerged (p=0.98, paired *t-test*) and did not differ from SD_{syll} at the end of development (p>0.98, paired *t-test*). At the same time, SD_{syll} for syllable B decreased strongly (p<0.003, paired *t-test*). This result indicates that exploratory variability was confined to the newly added syllable across birds.

We also compared variability during morning singing to afternoon variability. As will be shown in Chapter 4, morning variability of sub-syllabic structure tends to be higher than afternoon variability, suggesting that most exploration takes place during morning singing, which could be related to the “morning effect” (Derégnaucourt et al, 2005). Figure 2*A* shows examples from three birds. As in Fig. 2.2*D*, variability decreases across development in syllables A and B. Note that in some cases of A or B syllables the morning variability (traces marked with “*”) is consistently higher than afternoon variability (traces marked circles). This however was not the case across all birds in syllable B, while in syllable A the median variability was significantly higher in the morning than in the afternoon (Fig. 2.3*B*).

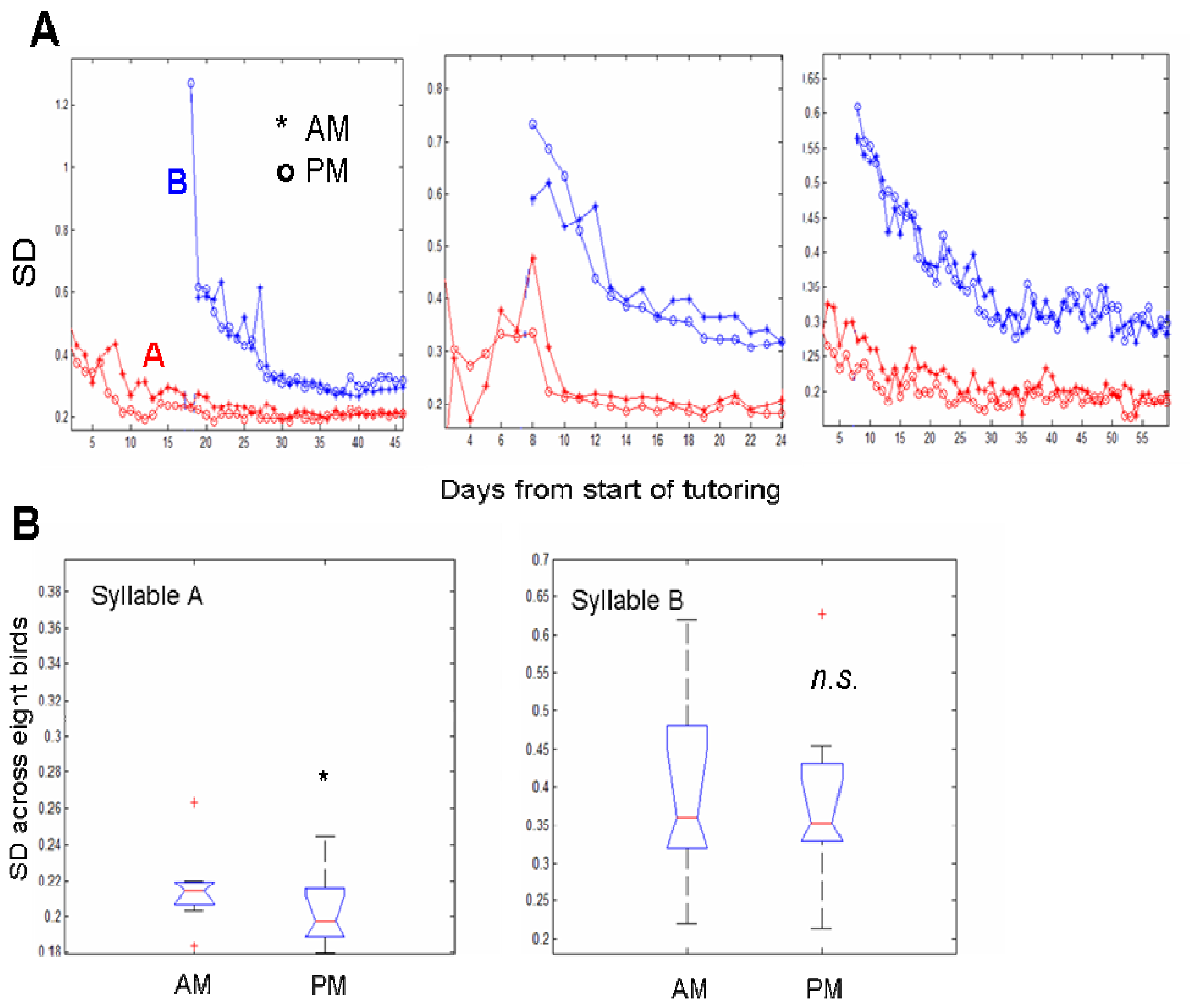

Figure 2.3 In some birds morning variability of syllables is higher than afternoon variability. Morning variability (SD across features) is shown with "*" symbol in ***A*** and afternoon variability with circles. Variability of syllable A is shown with red trances and variability of be with blue traces. Three examples of birds trained with sequential training procedure are shown. We did not find significant difference between morning variability and afternoon variability across all eight birds in syllable B. While syllable A is less variable than syllable B, its morning variability was higher than afternoon variability, across all eight birds, as shown in ***B*** (p<0.017, single-tailed *t-test*). We did not detect significant difference between morning and afternoon in syllable B. However, in the cases where there was a difference it was always in the direction of higher morning variability (as shown in

2.4 Conclusion and discussion

When new syllables are incorporated into a song bout, the exploratory variability required to learn the new model (target) is confined to them alone. This finding supports the hypothesis that, when faced with the conflict between consolidation and exploration during learning of continuous actions, the bird segments the singing action and adjusts vocal exploration level locally. We can conclude that learning trajectories of syllable A and syllable B are separate as they consolidate independently one from another. That is, the song can be learned in a piecemeal fashion.

What is the source of exploratory variability? As noted in Chapter 1, the temporary silencing of AFP ("learning pathway") by injecting TTX into LMAN promptly reduce variability between syllabic renditions to that of a fully stereotyped adult song (Olveczky et al, 2005). Similar results (elimination of variability between syllabic renditions) were obtained even in syllables that were highly variable (Andalman and Fee, 2009). It would be interesting in this context to perform the sequential training procedure in birds where LMAN could be inactivated. Given the above mentioned results the prediction would be that both, syllable A and the new syllable B would become stereotyped. A similar experiment could be carried out using directed vs. undirected singing. As noted in Chapter 1, during directed singing the syllables become much less variable (Hessler and Doupe 1999 a,b) and it has been shown that this difference in variability can be used to study the role of AFP in learning (Kao, 2005).

We took advantage of the “morning effect” to see if there is a difference in the diurnal variability oscillations associated with the effect, between syllables A and B. We predicted that if, as we will suggest in Chapter 4, most exploration takes place in the morning, the oscillations should be greater for the newly added syllable B. We observed a small, but significant difference between morning and afternoon variability in syllable A (p<0.017, single-tailed *t-test*) However, we did not observe any significant difference between morning and afternoon variability in syllable B It should be noted that in all birds where the difference in morning vs. afternoon variability was significant (the “morning “effect” was there) it was always the morning variability that was higher. However, as we shall see in Chapter 4, this is a misleading finding since variability at the syllable level is often a compound outcome of the means and variances of intra-syllabic events.

It is possible that while an overwhelming portion of variability could be explained by the active injection of noise to the motor pathway by the AFP (Olveczky et al, 2005) there is a second source of variability caused by differential sensitivity of the motor pathway to the AFP input. According to this view the morning variability could be higher because the motor pathway becomes more responsive (sensitive) to the noise injected by the AFP. This explanation would also be consistent with the view that morning song is more plastic than the afternoon song (Derégnaucourt et al, 2005).

Chapter 3: Regulation of vocal exploration during continuous actions

Most of the data used in this chapter was contributed by Dina Lipkind.

3.1 Background and Rationale

In Chapter 2 we show that exploratory variability can be adaptive such that a bird can confine the exploration only to the syllable that needs to improve most. The syllable that is already mastered does not become more variable even when the new variable syllable is added. We concluded that this can allow for a piecemeal learning of the song.

But what are the smallest units and the natural time scales of vocal exploration? One possibility is that these are the song syllables, which are somewhat discrete units of song production. Our null hypothesis is that vocal exploration is globally applied across the entire syllable. We will test if similar piecemeal learning strategy can take place in smaller units during the continuous action of the syllable. To answer this question we examined song learning at the sub-syllabic level, using song development data from six birds trained with the altered-target training procedure (as described in Chapter 2) and two birds trained with a single song model.

While quantifying the variability across renditions of whole syllables was relatively straight forward, getting such measures of variability of structure within a syllable was difficult because intra-syllabic structure (notes) are not easy to segment, and are not developmentally stable.

We therefore had to develop methods for robustly identifying significant time events during the continuous syllable, and for measuring the exploratory variability of those *intra-syllabic events*. For this reason Chapter 3 is dedicated to presenting these new methods and their application.

3.2 Methods

3.2.1 Training regimes and animals

The development of inta-syllabic events was measured in eight birds: Six of those birds were trained with alternate training procedure (see Methods in Chapter 2 for description of the procedure and Fig. 2.2 *B* for the song model), and in each of those birds we analyzed the first syllable (A) they were trained with. Syllable A has complex structure and having six repeats of the learning of the same syllable allowed us to compare the learning of specific intra-syllabic song elements across birds. To make sure that our findings are not limited to this training regime or to the particular choice of syllable, we analyzed another syllable from two birds that were trained with a single song model (Fig. 1.1). For the analysis of intra-syllabic events we used the one complex syllable from those two birds. All the findings described in this dissertation apply to both groups, and since we have seen no apparent differences in results, we pooled the data and present those eight birds as a single group.

3.2.2 Syllables selected and analytic approach

For each bird, we selected for analysis one complex syllable with at least three distinct song elements (but typically 5-6 song elements) at the endpoint of song development and tracked occurrences of local minima in Wiener entropy within each syllable across development.

In previous studies (Du and Troyer, 2006), complex song syllables were often segmented to smaller units called notes. Although segmenting syllables to notes is possible in adult zebra finches, we find this task much more difficult during song development. A somewhat easier task is of detecting significant time events within a syllable. The advantage of this approach is that it does not require detection of boundaries (which are often blur in developing songs) but only center of events. Further, if, due to noise, our probability to detect such events is not particularly high, say 50%, a large enough sample of syllables should allow us to detect those events quite easily. In contrast, trying to segment a syllable to notes with 50% noise level in detecting boundaries would have resulted in highly variable segments.

One of the most robust features of the developing song is Wiener entropy, and its minima indicate a local peak in tonality (highest local concentration of spectral energy). To detect significant time events within syllables, we measured Wiener entropy minima for all occurrences of a syllable type for each day of development (roughly 600-10,000 renditions per day, per bird).

3.2.3 Smoothing procedure

For each rendition of the syllable type we smoothed the Wiener entropy time course using 30 ms Hanning window using the Matlab 7 *hann* function and then automatically detected the local minima in the smoothed curve (usually 3-6 per syllable). These local minima often consistently repeated across syllabic renditions to form clearly visible clusters (see Figure 3.3) that we call “intra syllabic events” (see Fig. 3.3). The 30 ms size of the Hanning window was determined empirically in a subset of our data to optimize the detection of distinct intra-syllabic events. Experimenting with different windows sizes, we found no or little impact of the window size on the qualitative results: the same clusters (intra-syllabic events) were identified over a range of different window sizes. Further, in the vast majority of the cases we were able to trace intra syllabic events over prolonged developmental epochs.

3.2.4 Time histograms of significant events

In order to identify intra-syllabic significant events (consistently identifiable events of Wiener entropy minima within a syllable) we aligned samples of 100 syllables of the same type and computed the histograms of time positions of Wiener entropy minima. Those histograms had sharp peaks that could be easily detected at the time-positions of highest density, which we call intra-syllabic events. The minima of the histograms were used to delineate the outer limits of each intra-syllabic event. These limits (separations between different intra-syllabic events) were set 5 milliseconds in each direction from the histogram minima.

3.2.5 Detecting song elements within a syllable

Much of our analysis was based on detecting robust song elements within syllables during song development. As mentioned above, our goal is to analyze the role of exploratory variability within a syllable. We encountered two problems arising from measuring the variability across renditions of continuous actions: alignment and segmentation. For example, to estimate the variability, one could align renditions of syllables according to the points where a particular acoustic feature crosses a threshold, as shown in Figure 3.1*A*.

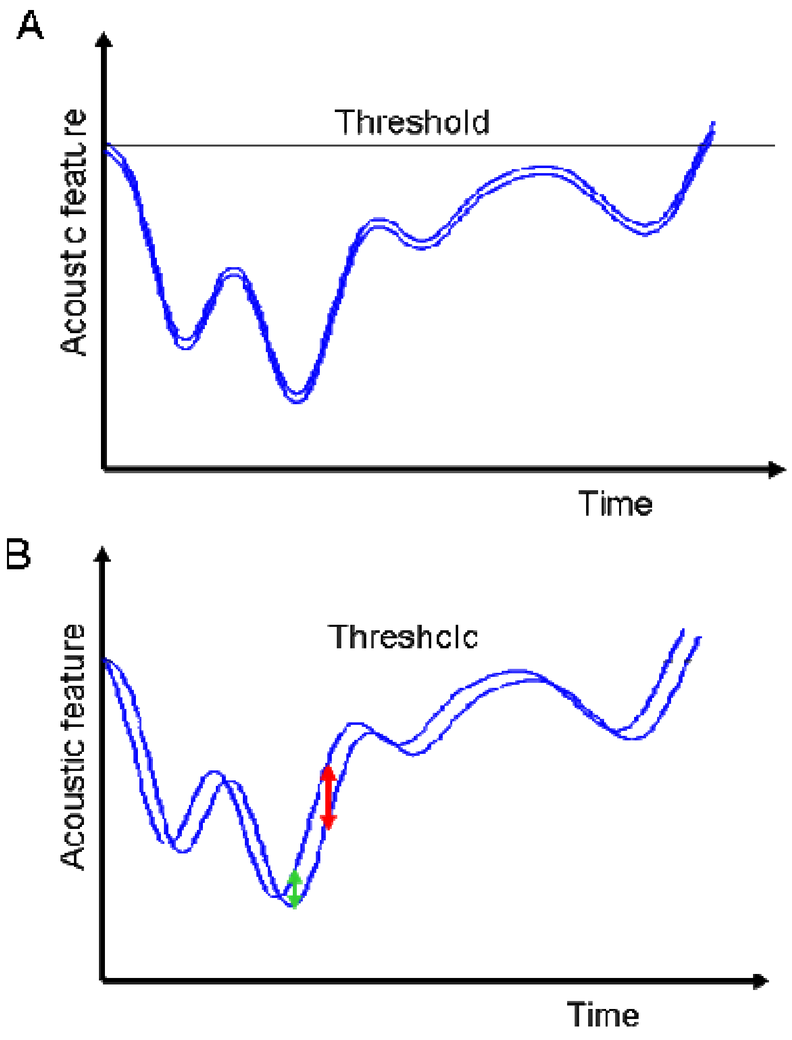

Figure 3.1 Misalignment of syllables can result in biased measurements of variability across renditions such that the slope and variability estimates will be positively correlated. See text.

Now one could measure variability across renditions as standard deviation, for each millisecond of the syllable. With perfect alignments this simple method should produce reliable measures of variability at each time point (millisecond) within a syllable. But the problem appears when there are slight misalignments between syllabic renditions. Note that such “misalignments” can result from bird’s own temporal imprecision. The problem is illustrated in Figure 3.1*B*.

Note in Fig. 3.1*B*, that when the slope of syllabic features is low, the measured difference between the two renditions (and thus SD) is also low (green arrow) while in time-positions located at high slopes, the measured difference between renditions is relatively high (red arrows). Therefore our measure of variability between syllabic renditions will trivially depend on the location within a syllable. Another way to see the problem with this method is this: we are comparing the values of acoustic features between renditions of syllables at a particular *time* within a syllable. For example, we compare what a bird did at millisecond 20, in one syllabic rendition, to what he did at millisecond 20, in another rendition. But millisecond 20 does not necessarily correspond to a particular (consistently identifiable) acoustic event within a syllable.

In order to circumvent this problem of comparing syllabic renditions at each millisecond, we identified events within a syllable that were consistently present in each rendition.

We found, empirically, that Wiener entropy is the acoustic feature with which we could identify such consistently present events, early in development after the first appearance of the syllable (we will discuss the time of appearance of these events later). We experimented with using other acoustic features (amplitude, mean frequency and pitch) for the purpose of identifying significant time events (traceable across development) but Wiener entropy produced the best results. We therefore used Wiener entropy as a master feature to identify time-positions of the intra-syllabic events. From these time-positions we extracted the values of other acoustic features as well.

Wiener entropy is a measure of spectral noise (it estimates the width of the power spectrum). Instead of linear scale (zero entropy = order, one = disorder) we use a logarithmic scale to increase the dynamic range, and zero (log(1)) stands for disorder, i.e., white noise, and negative values for order (harmonic stacks and pure tones, Tchernichovski et al., 2000). A Wiener entropy local minimum represents a moment within the syllable where the concentration of spectral energy reaches a local peak – this corresponds to moments where harmonic stacks or pure tone notes are most clearly defined, as shown below (Fig. 3.2):

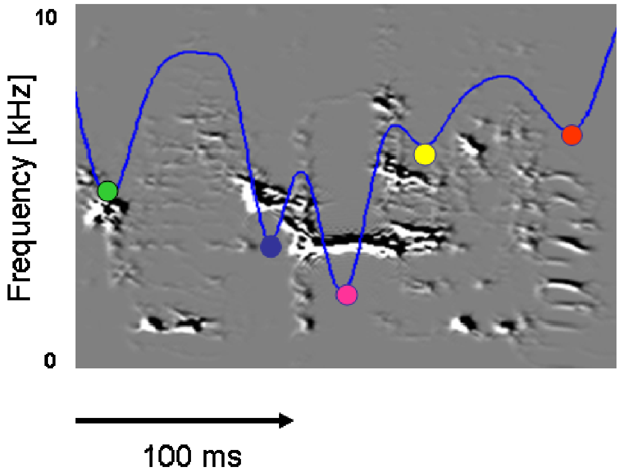

Figure 3.2 Detection of significant intra-syllabic events. Wiener entropy was used as the acoustic feature in order to identify significant intra-syllabic events. The blue trace, overlaying the spectrogram of a syllable, is the smoothened Wiener entropy. The local minima were selected as significant intra-syllabic events (colored circles) if their time positions were consistent across syllables. Note that the intra-syllabic events thus identified capture the points of high local spectral density (or harmonic sounds), with the middle event (pink circle) as the global Wiener entropy minimum.

In Fig 3.2 above note five detected events within a syllable indicated by colored circles. The blue trace represents the smoothened Wiener entropy (see Methods in this chapter for smoothing procedure). If these events consistently repeat across syllabic renditions, we call them significant intra-syllabic events. In the next sections we show how these intra-syllabic events can be tracked.

3.2.6 Tracking intra-syllabic events

In previous section we showed how we detected significant events within a syllable that could be used to compute variability across syllabic renditions, circumventing the alignment problem. For these events to be useful they have to be present consistently across syllabic renditions, e.g. we should be able to track them over the time of syllabic development.

Figure 3.3 demonstrates the identification of intra-syllabic events as consistently occurring events within a syllable.

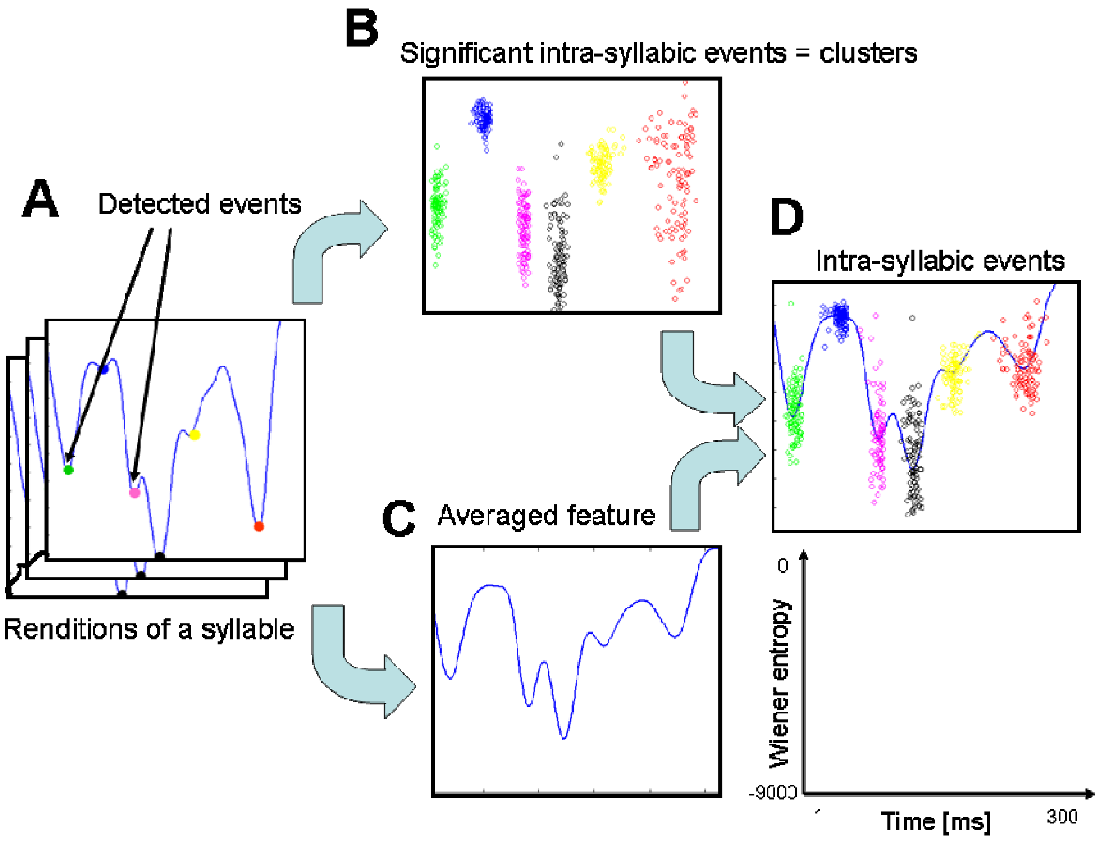

Figure 3.3 Procedure for identifying of intra-syllabic events. ***A***: smoothened Wiener entropy (blue trace) of one syllabic rendition. Six significant events (Wiener entropy minima) were identified: colored circles. ***B***: Clusters of significant events form after several syllabic renditions. These clusters we call "intra-syllabic events". ***C***: Averaged (across renditions) shape of Wiener entropy. ***D***: Averaged Wiener entropy and intra-syllabic events overlaid.

First Wiener entropy local minima were detected, as described in the previous section (Fig. 3.3 *A*). After several renditions of a syllable, these local minima were detected as clusters (Fig. 3.3 *B*). Note that the clusters are well separated from each other in time, so a time histogram could be used to identify them automatically. We called thus identified clusters *intra-syllabic events*.

Some intra-syllabic events are more “subtle” then others, meaning that they do not necessarily appear in each rendition of a syllable. For example, in Figure 3.3*A* the blue and yellow minima in Wiener entropy detected are less pronounced than other minima and are not detected in every rendition of the syllable. As a result the total number of detected minima over a day varies among different intra-syllabic events. For this reason we computed all statistics of intra-syllabic events using a fixed sample size of each intra-syllabic event (i.e., using the same running widow size as presented in section 3.3.2.). Since we record and analyze the entire song production of each bird, keeping sample size equal across events has only a negligible effect of having data epochs with slightly different boundaries for the analysis of different intra-syllabic events.

In summary, intra-syllabic events are Wiener entropy minima that clearly and consistently appeared in a recognizable timing within a syllable.

3.3 Results

3.3.1 Tracking of intra-syllabic events across syllable development

Figure 3.4 demonstrates the tracking of intra-syllabic events across syllabic development in one bird.

On the third day after the onset of training, a syllable first appeared as a district type (identifiable cluster in feature space such as in Fig. 3.4*C*). Three to seven days later, events appeared that seem to fall into clusters as described above (Fig. 3.4*B*).

Figure 3.4*A* shows the approximate location of the intra-syllabic events on the sonograms. Notice that the almost pure tone event (black color) appeared after much later than other events.

This pattern of intra-syllabic event appearance was observed in all eight birds: syllable types appeared 3.2±2.1 days (means and SD) after the onset of training while intra-syllabic events appeared 7.6±3.1 days after the training onset. This delay suggests a hierarchical process, where coarse structures (at the syllabic level) consolidate prior to the fine structures (at the sub-syllabic level).

In practice, once an *intra-syllabic event* could be recognized, we were usually able to track it continuously and automatically thereafter (using time histograms as described above). In some cases such as bifurcations we had to identify clusters by visual inspection.

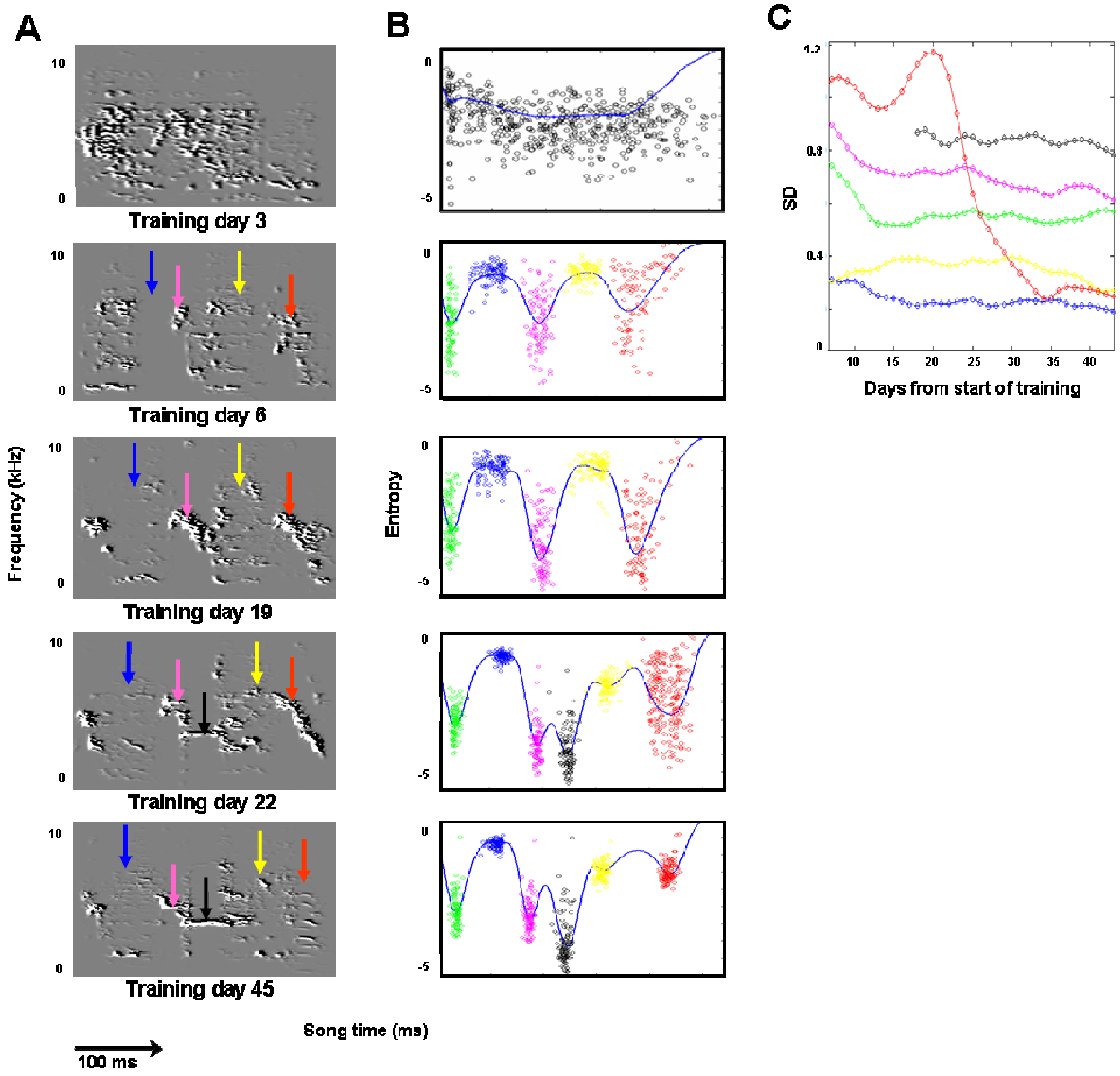

Figure 3.4 Time courses of changes in variability within a syllable. ***A,*** Sonogram of a developing syllable from its first appearance (day 3) until its consolidation (day 45). Arrows point to approximate locations of the intra-syllabic events identified. Note that on Day 3 we could not yet identify any intra-syllabic events. ***B,*** Identified intra-syllabic events can be traced across trials. The blue curve represents mean Wiener entropy traces averaged across 50 syllables. Overall, the intra-syllabic events become less variable as the clusters get smaller. (The variability of the red cluster, however, seems to increase on day 22.) Note that at day 3 we can identify syllabic types but Wiener entropy minima (black circles) do not form intra-syllabic events (clusters) until day 6. ***C,*** Developmental time courses of variability ($SD_{intrasyll}$) for all identified intra-syllabic events. Variability decreases across all intra-syllabic events but this decrease is asynchronous.

With the method of tracking intra-syllabic events at hand, we can now analyze each event separately across the course of syllabic development; we can measure variability of each intra-syllabic event, independently of alignment.

3.3.2 Statistics of intra-syllabic events

In the previous section we show how intra-syllabic events can be identified tracked independently of the alignment. Because we analyzed complex syllables, with at least three intra-syllabic events (typically 4-5 events), this method also allows us to segment the syllables. This allows us to compute the statistics for different parts of the syllable across renditions.

The two main statistics that we used were mean and standard deviation. The former was used to determine the feature magnitude of a part of a syllable and the latter was used as a measure of exploratory variability for each part. As noted in previous sections Wiener entropy was the main acoustic feature with which intra-syllabic events were identified. Values of other acoustic features were then identified at the time-positions of these intra-syllabic events.

Mean values and standard deviations across 50 renditions were computed for all features of intra-syllabic events. We will refer to the variability of intra-syllabic events as $SD_{intra\text{-}syll}$. In order to compute $SD_{intra\text{-}syll}$ we first removed the trend (the shift of the mean) which results from the natural progression of learning. This procedure produced a few hundreds of mean and variability estimates per day. To obtain daily estimates of variability for each day we pooled the standard deviations, computed as described above, for whole-syllable features. To avoid the strong oscillations in song structure during mornings (Derégnaucourt et al., 2005), we pooled $SD_{intra\text{-}syll}$ values from the latest third of daily renditions only (which corresponded approximately to evening singing) to obtain stable daily estimates. We will discuss these daily oscillations in Chapter 4.

Figure 3.4*C* is showing time courses of daily values of variability ($SD_{intra\text{-}syll}$) for six intra-syllabic events. Note that the variability of all six events is decreasing but this decrement is not synchronous: the variability of the red intra-syllabic event, for example, increased approximately from day 15 to day 25 after the start of training and decreased thereafter (this is also evident from the size of red cluster in Fig.3.4*B*), while the variability of most other intra-syllabic events in Fig. 3.4*C* decreased before that period.

We wondered if such asynchronous decrements of variability could indicate the presence of adaptive exploration (regulation of exploratory variability) within a syllable. Because we did not have the same advantage of alternate training procedure to manipulate learning of intra-syllabic events (as we did with whole syllables, see Chapter 2) we had to look at how close intra-syllabic events were to the target (the song model). We turn to this issue in Chapter 4.

Chapter 4: How the developing song is compared to the song model

4.1 Background and Rational

In Chapter 3 we discussed the difficulties of quantifying the variability across renditions of syllables. We introduced the method for identifying and tracking of the intra-syllabic events which ultimately can be used to compute the unbiased estimates of exploratory variability on short time scales (short lasting events inside of a syllable).

In the present chapter we turn to our main question: do birds regulate exploratory variability on a short time scale, within a syllable (which is a true continuous action). As discussed in the Introduction birds posses the brain pathway responsible for song learning (AFP), which can induce fine time-scale changes to the song performance (Kao et al., 2005) and this can be used for exploration of acoustic space. So while the “hardware” seems to be in place for such localized application of exploratory variability we do not know if birds actually use it this way.

Another issue that we discuss in this chapter is the question of diurnal oscillations of syllabic structure and the role of exploratory variability in the morning singing (when variability is higher).

With our method for measuring the exploratory variability of intra-syllabic events at hand we now look at how variability correlates with acoustic error. But first let us define “local error” and “global error”.

4.2 Methods and definitions

4.2.1 Local acoustic error of intra-syllabic events

Local error was obtained by computing the Euclidean distances for each intra-syllabic event relative to the corresponding events in the target. Distances were computed for each day using the daily means of acoustic features. These daily measures provide a time course of the local errors over development. This time course was correlated with the variability of intra-syllabic events which we call the *local correlation* (r_{local}).

4.2.2 Global acoustic error of intra-syllabic events

To estimate the global errors for intra-syllabic events, we computed the mean for each syllable of the local acoustic errors across intra-syllabic events. Again, these values are obtained for each day and correlated with the variability of intra-syllabic events which we call now the *global correlation* (r_{global}). Notice that we obtained these global correlations for each intra-syllabic event separately.

4.3 Results

4.3.1 Correlating exploratory variability with acoustic error

To visualize an example of the detailed relationship between variability and distance from target we plotted them against each other as shown for one syllable in Figure 4.1*A*, *B*. Distances of Wiener entropy and mean frequency were computed in reference to the end-point (Fig. 4.1*A*) and also in reference to the song model (Fig. 4.1*B*). As shown, in all intra-syllabic events, variability decreased across the time of development, but at different rates. For most events there was an association between the distance from the song model (or from the end-point) and the level of variability. For example, looking at Fig. 4.1*B*, for events 1 and 6, the initial variability was high in both, but in event 1 variability and distance from target decreased substantially in approximately five days, whereas in event 6 variability and distances both remained at high levels for about 20 days. Another interesting case is event 3, which bifurcated after 10 days into events 4 and 5. Note that event 4 was initially closer to the target and also of lower variability than event 5.

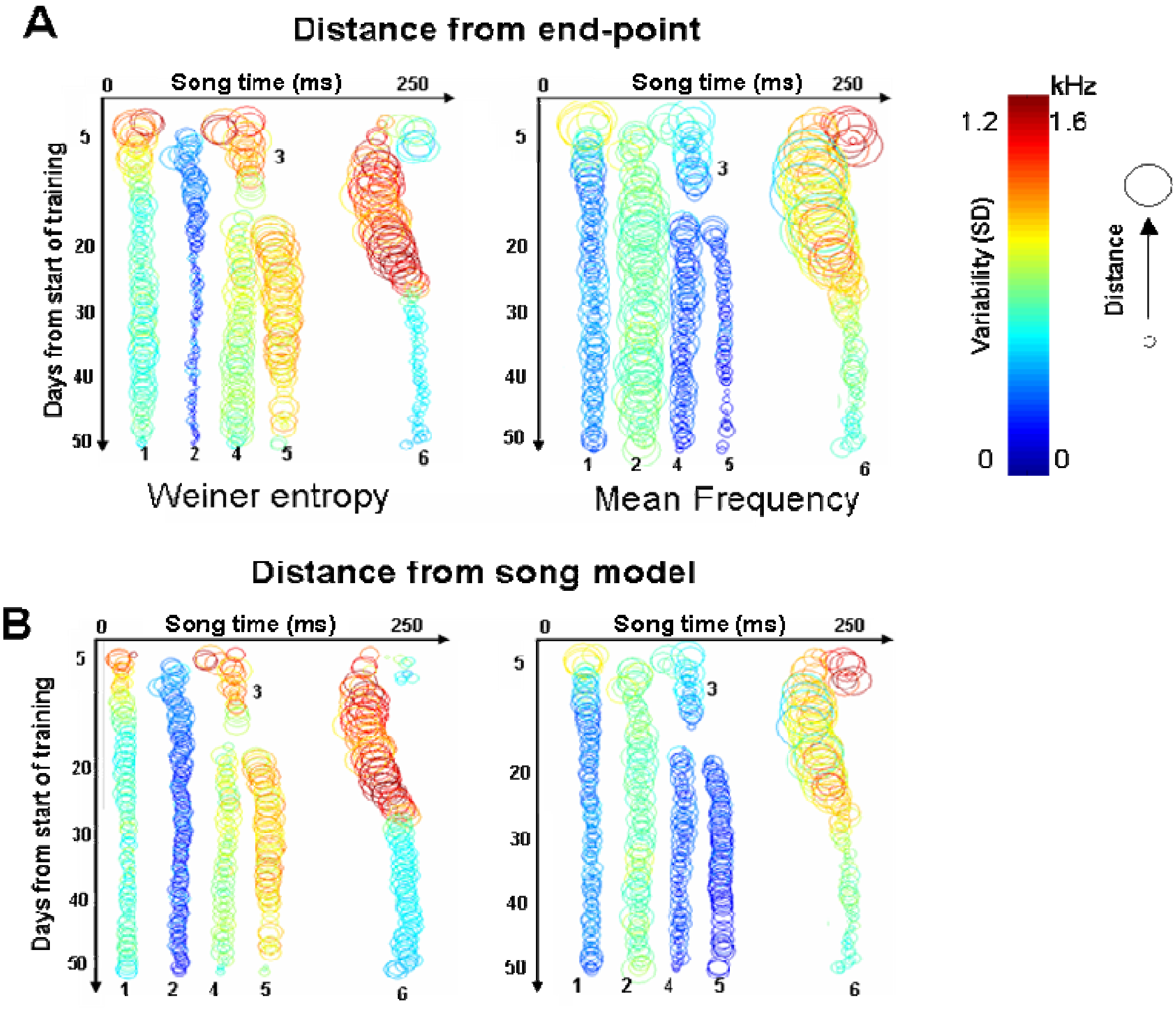

Figure 4.1 Variability decreases when intra-syllabic events reach their targets. ***A,*** Across development, variability of significant events decreases near their local end-points. Color of circles corresponds to $SD_{intrasyll}$ and size represents the Euclidian distance of intra-syllabic events from their end-points. ***B,*** The same variability data as in ***A*** is presented with the Euclidian distance from the corresponding intra-syllabic event in the song model.

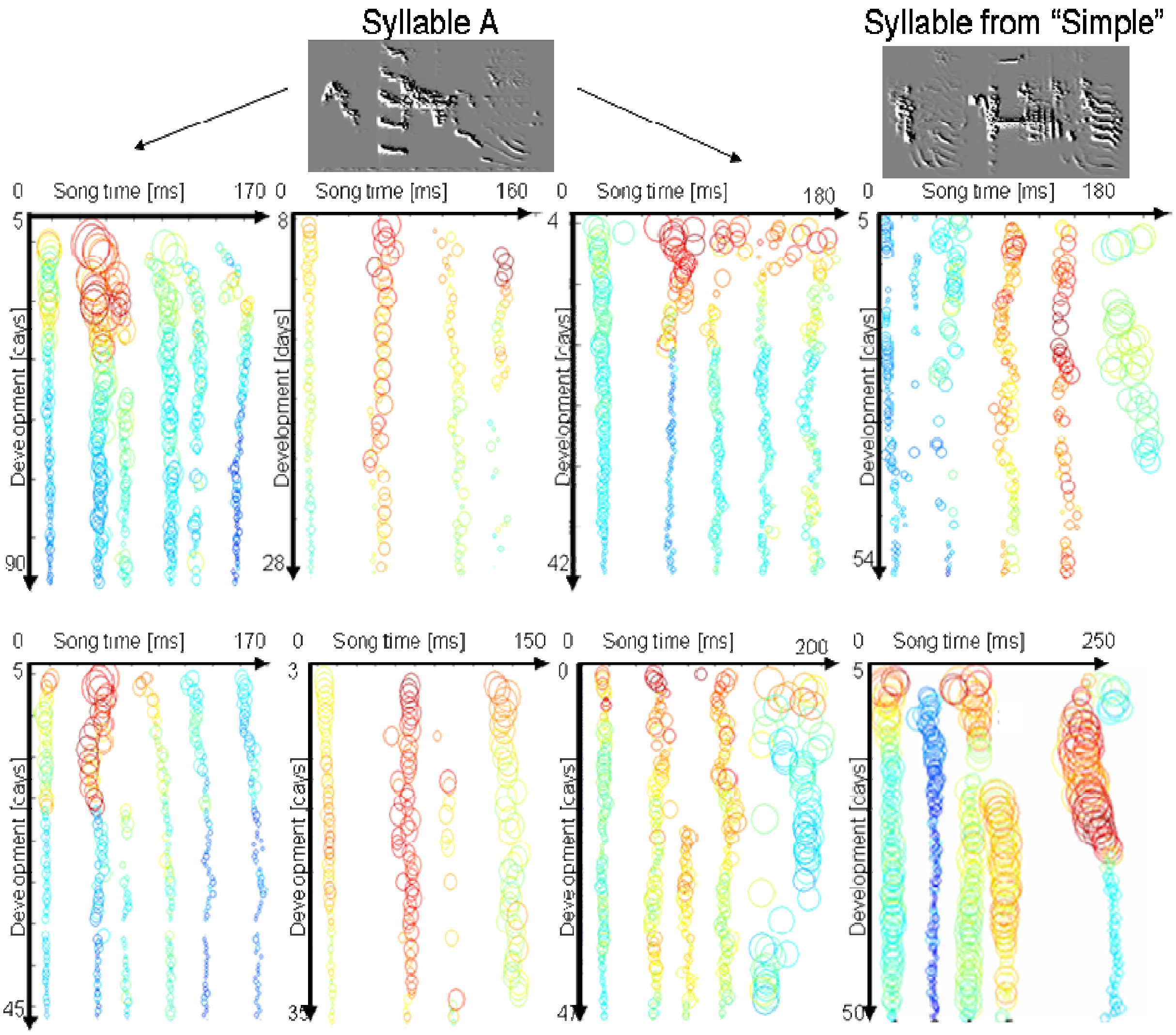

Figure 4.2 Exploratory variability and distance from target across birds. In most intra-syllabic events, across all eight birds studied the magnitude of variability $SD_{intrasyll}$ and the distance from the target are correlated. As in Fig 4.1 color of circles corresponds to $SD_{intrasyll}$ and size represents the Euclidian distance of intra-syllabic events from their corresponding events in the target song. The feature shown is Wiener entropy. Six birds were trained with sequential training (see section 2.2.4) and two birds were trained with a single song model (“Simple”). For both models the most complex syllables was used in the analysis.

Figure 4.2 shows the how exploratory variability relates to the distance from the target across all eight birds used in this study.

4.3.2 Variability correlates better with local error than with global error

We wondered if the magnitude of vocal exploration differs across intra-syllabic events and if so, whether different time courses of variability might mirror differences in the learning pace of different intra-syllabic events, i.e. whether decrease in acoustic error drives adaptive decrease in variability separately for each event. To test for this we computed the Euclidean distance of intra-syllabic events from the corresponding events in the target (*local error*). We then examined if this local error can explain the local variability ($SD_{intrasyll}$) better than the *global error* over the entire syllable. Note that the global error (being the average error of intra-syllabic events in the syllable) provides a more stable estimate compared to individual local errors, and therefore by default should provide slightly better correlations. We first computed the time course of global error across development, and correlated it with the time course of variability for each intra-syllabic event ($SD_{intrasyll}$). This estimate, r_{global}, was then compared to the correlation between the local errors and local variability, r_{local}.

To test if exploratory variability correlates locally within a syllable, we analyzed eight complex syllables, one from each bird trained with either sequential training procedure (six birds) or a single song model (two birds, see Method sections in Chapter 3, 3.2.1 & 3.2.2). The eight syllables selected (always the most complex syllable in a song in each bird) typically had 4-6 intra-syllabic events. The relationships between distance from the target and variability ($SD_{intrasyll}$) intra-syllabic events, across all eight birds are shown in Figure 4.2.

We computed r_{local} and r_{global} for each intra-syllabic event in each of the eight syllables, which produced 41 values for r_{local} and r_{global} .

Figure 4.3 shows plots of r_{local} vs. r_{global}, for Wiener entropy, mean frequency and time position. The data include all identified intra-syllabic events across all birds (*n*=41 intra-syllabic events from 8 birds). Measures of intra-syllabic events within a syllable cannot be considered as independent samples. Therefore, for the purpose of statistical testing, we computed median correlations (r_{local} and r_{global}) across intra-syllabic events for each bird, to obtain a single statistical estimate per bird. We restricted this analysis to differences from the song model (differences from the endpoint were only inspected qualitatively). Variability was significantly more correlated with local error than with global error in all song features tested: Wiener entropy (p= 0.015) mean frequency (p= 0.04) and time positions (p= 0.044, paired, two-tailed *t-test*; n=8 birds). Results in reference to end-points and song model were similar, but there were more cases of negative correlations when the reference was a song model. Investigating these showed that the bird had reached the model values, but then continued with vocal changes further away from the model, perhaps purposely diverging from the model.

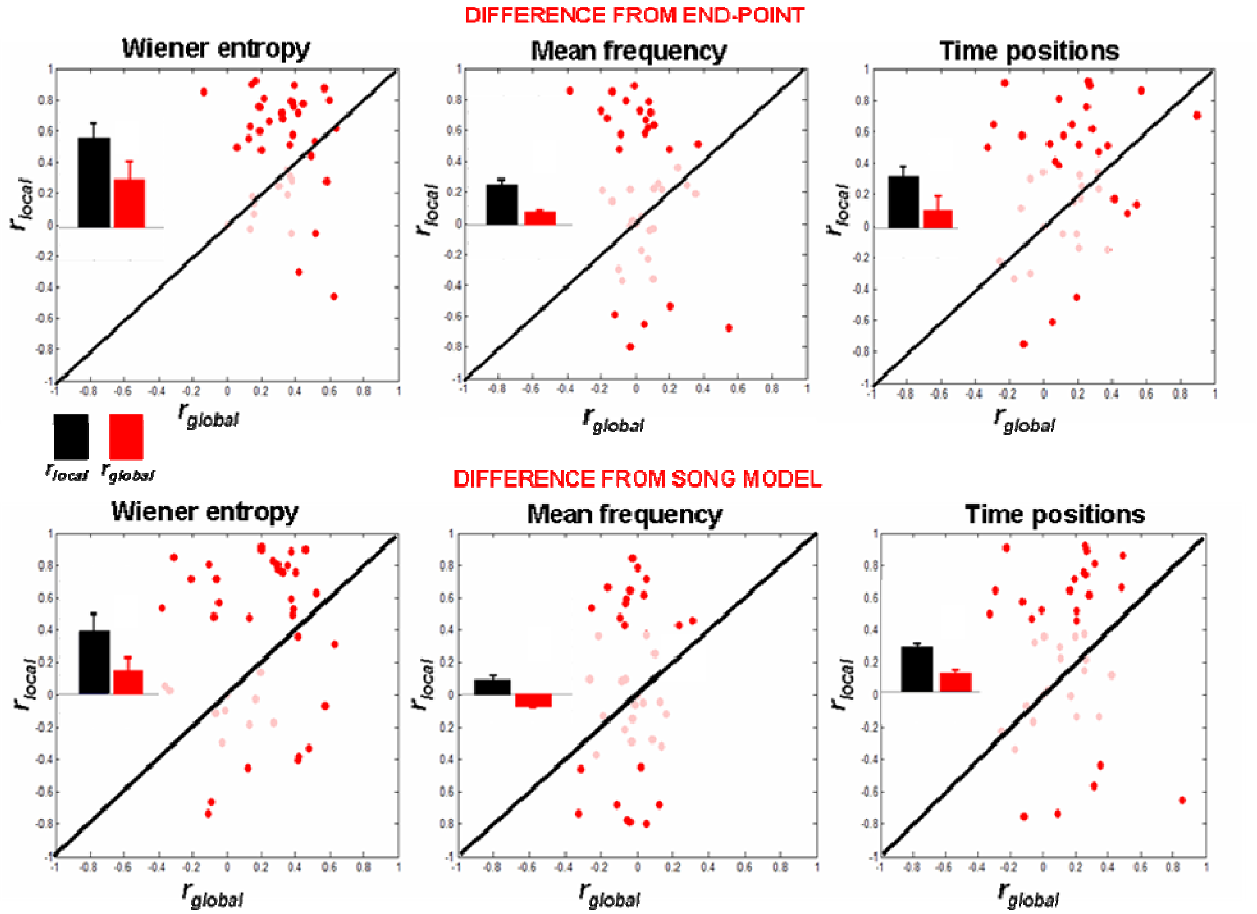

Figure 4.3 Variability of intra-syllabic events is better correlated with local than with global error. We correlated time courses of variability ($SD_{intrasyll}$) of intra-syllabic events with their local error to obtain r_{local} and with global error to obtain r_{global}. The errors were computed for Euclidian distances from both end-point (top) and the song model (bottom). Local and global correlates (r_{local} and r_{global}) were computed for all intra-syllabic events, across 8 birds, separately for each feature. Local correlates are significantly greater than global correlates for all three song features (see text). For visualization, we denoted non-significant correlations by pink circles. The bars present mean and s.e.m for correlations with local (black) and global (red) correlations.

4.3.3 Exploratory variability changes simultaneously with local error

We next examined if changes in variability ($SD_{intrasyll}$) and local error occurred *simultaneously* within the one-day time resolution of our study. To test for this, we computed correlations introducing lags between error and variability (cross-correlations). We computed lags for all three song features in all intra-syllabic events (across all eight birds). We then computed histograms of lags for all three features (y-axis represents the number of events and x axis the lag). Thus we obtained three histograms, one for each song feature. Next we combined the three histograms (added the frequencies) and obtained a single histogram representing all three song features, as shown in Figure 4.4. There was no significant deviation of correlations from lag=0 (p=0.512, *paired t-test*). This result indicates that changes of variability and local error are indeed simultaneous within the one-day time resolution analyzed.

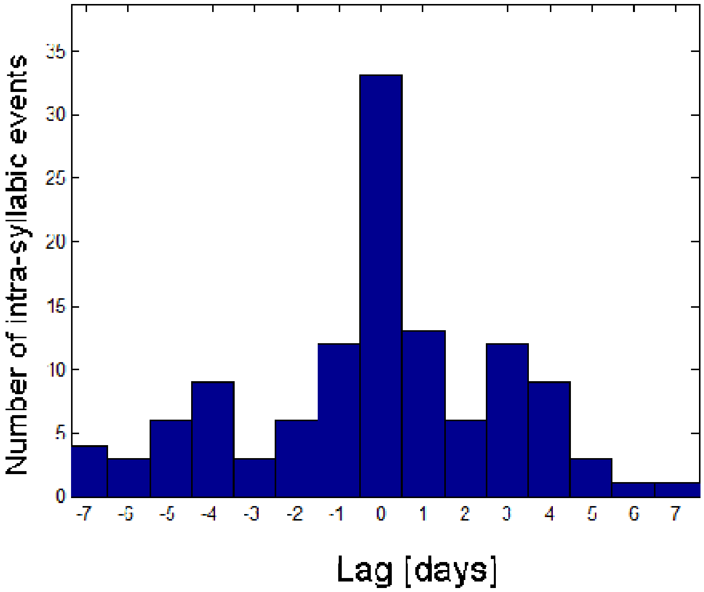

Figure 4. 4 Exploratory variability changes simultaneously with local error. We computed cross-correlations of local error and exploratory variability ($SD_{intrasyll}$) for all intra-syllabic events across all eight birds studied. Histograms of lags were computed for each of the three song features. In every feature the highest number of events had lag of zero. The histogram presented here is the cumulative histogram for all three features (sum of the three histograms).

An alternative explanation to our findings would be that perhaps certain song elements are both easier to learn and faster to stabilize (regardless of learning). To test for this we compared the learning of four equivalent intra-syllabic events in six birds that were trained with the same target syllable. We assessed the learning speed of each intra-syllabic event by computing the overall change in distance from the song model, and calculating the developmental time when the event reached half of this distance; *ANOVA* of these learned speed estimates showed no significant difference between the intra-syllabic events (p=0.124; F=2.25; across n=6 birds). Also note in Figure 4.2 that there seems to be obvious order at which intra-syllabic events consolidate in the six birds trained with syllable A (the sizes of the circles represent the distance from the target). Therefore, different birds learned similar intra-syllabic events in different relative rates, and the only correlation that holds is the one between the local error and variability.

Taken together these results indicate that the changes in local variability are best explained by the local error from the model, suggesting that birds can evaluate the error locally, at short time scales of no more than 20-50 ms, and maintain high exploratory variability primarily in those intra-syllabic events where local error is high. It is particularly interesting that variability not only of spectral features but also of time positions was better correlated with local error (of timing) than with global timing error, which suggests that time-jitter of intra-syllabic events is locally gated within a developing syllable.

4.3.4 Diurnal oscillations in variability

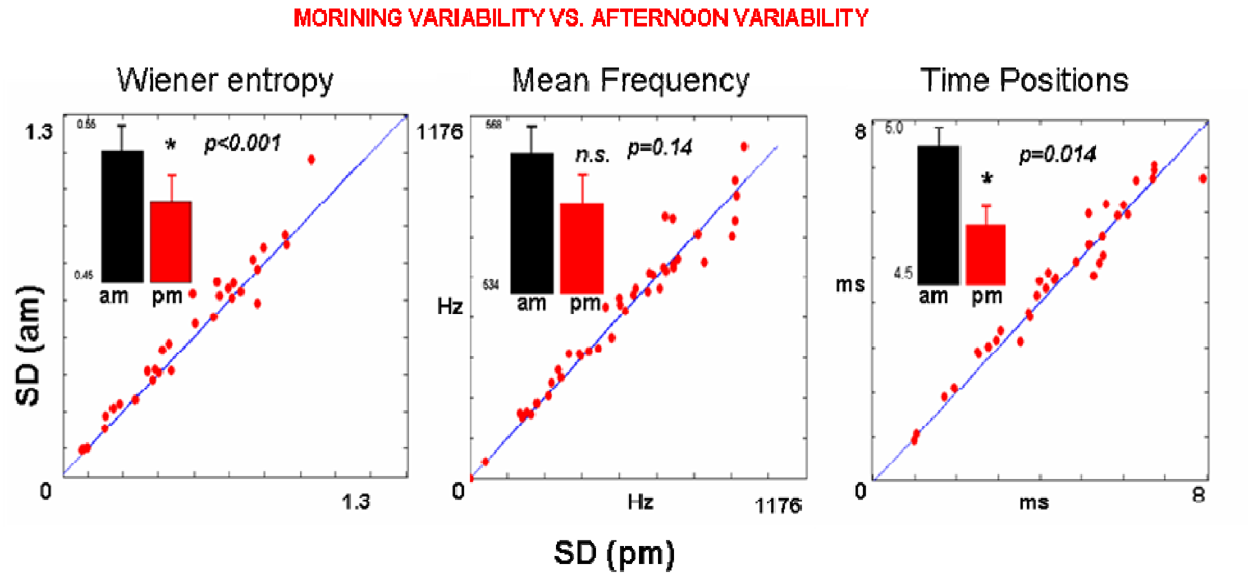

Figure 4. 5 Variability of intra-syllabic events tends to be higher in the morning than later in the day. Time courses of variability ($SD_{intrasyll}$) were computed for each intra-syllabic event from samples taken in the morning and was compared to total variability time course taken from late afternoon samples. In two song features (Wiener entropy and time positions) variability was significantly higher in the morning than in the afternoon in ($p<0.001$ for Wiener entropy and $p=0.02$ for time positions). For mean frequency the difference between morning and afternoon variability was not significant ($p=0.14$).

In the analyses thus far, with daily units taken from the afternoon and evening songs, developmental time courses of error and variability were often monotonic. However, previous investigations have shown that during periods of rapid learning, there are strong diurnal oscillations in song structure, such that the morning song is less structured and less similar to the song model (Derégnaucourt et al., 2005). We would predict that if a bird can adaptively change the magnitude of variability, the morning song, which is less similar to the model, should be more variable than the afternoon song. But a previous study (Miller et al., 2010) showed that syllables become more, rather than less, variable after daily practice (comparing morning to afternoon songs). We therefore performed a similar analysis, but instead of examining variability at the syllable level, we computed variability of intra-syllabic events.

In order to assess diurnal effect we repeated the computation of variability as above, except that we segregated the data into the first and last thirds of the day (corresponding approximately to morning versus evening renditions). Note that estimates of variability obtained by our method do not depend on the alignment of syllables nor do they depend on the modulation of syllabic features. Highly structured syllables tend to be more modulated while in less structured syllables acoustic features are more "flat," which could result in biased estimates of variability because higher diversity of vocal states (typical of highly developed syllables) may produce higher estimates of differences between syllabic renditions, which in turn may result in higher estimates of variability (across renditions). Identifying intra-syllabic events circumvents this problem.

As shown in Figure 4.5, variability across all intra-syllabic events is significantly higher in the mornings for Wiener entropy and time-positions, which is consistent with the hypothesis that young birds can increase or decrease variability adaptively.

4.4 Conclusion and discussion

Putting problems with articulatory dynamics (such co-articulation) aside, partitioning the continuous action could reduce the complexity of the song learning task. Partitioning the learning of a continuous action into discrete tasks is sometimes considered suboptimal (Doya, 2000). At one extreme, the action can be segmented very coarsely and error can then be computed for each segment. At the extreme coarseness of segmentation (single segment) this would mean computing a global error. At the other extreme, the action can be segmented into a very high number of partitions, which would necessitate the computation of many local errors. This would require large memory capacity and many learning trials (Doya, 2000).

In zebra finch song, however, the overall number of distinct song elements (syllables and notes) ranges only from 5 to 20 and so the scale of partitioning does not need to be too fine. To illustrate the advantage of partitioning consider, for example, the learning of 10 syllables, each with 10 possible vocal events (such as our intra-syllabic events). Assuming that the bird learns these vocal events one by one, and notes do not interfere with each other (no co-articulation), then the bird needs to learn 10 tasks of 10 states (100 possible states). However, if the error is only available globally, than the bird would have to select among 10^{10} possible states (regardless of the precision of the error computation).

4.4.1 Age dependent plasticity and the gain of exploratory variability

Zebra finches rapidly lose their song learning abilities with age and with learning as their song turns from variable to stereotyped (Morrison and Nottebohm, 1993; Boettiger and Doupe, 2001), but they nevertheless retain some plasticity into adulthood. The minor residual variability in song features across renditions, that persists even in "stereotyped" song syllables, is still accessible to reinforcement learning, and can be used to train birds to shift the fundamental frequency of a targeted vocal element within syllable up or down (Tumer and Brainard, 2007; Andalman and Fee, 2009). In these experiments a short burst of white noise was played to a singing bird as a negative reinforcement, at the precise time when a particular vocal event was detected in the bird's song. The negative reinforcement was removed if the bird either increased or decreased the fundamental frequency (to reach the "escape frequency" defined by the experimentalists). This training procedure resulted in consistent modification of the targeted vocal element. Interestingly, however, once the negative reinforcement training was finished, the fundamental frequency of the vocal element started to shift back to the original value. This suggests that the error (deviation from the original song model) is persistently reported after the song is modified. It seems that the memorized "template" does not change even if the song is modifiable in adult birds.

However, even prolonged training of adult birds which succeeded to shift their fundamental frequency did not induce any increments in the gain of variability in the targeted syllables (Tumer and Brainard, 2007). If the "escape frequency" that a bird had to reach (in the targeted vocal element) was too far from the original fundamental frequency, the animal never succeeded to modify the song and escape the negative reinforcement. This result proves that residual exploratory variability is indeed required for birds in order to locally modify their song but also implies that they can only do this in small steps (not exceeding the gain of residual variability). This also suggests that the developmental transition from high to low variability cannot be easily reversed. (For if it was easily reversible, then why would birds “choose” to not increase the exploratory variability of the targeted vocal element?)

These findings are reminiscent of the results by Knudsen and Knudsen, which show the differences in plasticity between young and old animals in terms of adaptive adjustment of auditory orienting behavior in response to displacing-prisms in barn owl (Knudsen and Knudsen, 1989). Juvenile owls can calibrate their auditory map to adjust for large angular errors in their visual field, but adult owls can only adapt to small errors and thus only learn in small steps. In the adult owl, the ability to adaptively rotate the auditory map is constrained by a narrowing integration window at the neuronal level. Analogously, in songbirds, as we have seen, it is the range of active vocal exploratory variability that constrains learning (Tumer and Brainard, 2007). Vocal exploration in juvenile birds provides a broader range of usable song elements than the adult song, due to stronger variability within a syllable and across syllabic renditions as well as stronger diurnal oscillation in song structure (Derégnaucourt et al., 2005).

Our interpretation of this result is that different parts of the song crystallize independently based, at least in part, on local error. This interpretation can explain the rapid changes in the gain of variability between syllables A and B documented in Chapter 2, as well as the lack of evidence that birds can easily increase the gain of local variability. This can be shown by an analogy:

Imagine walking through a corridor while looking outside via windows with variably sized openings: each time we cross a narrow window we are forced to look at the same image, but while crossing a wide-open window we have a range of images to choose from. Consequently, variability in the position of our eyes changes quickly as we walk by the windows, although the opening width of each window remains unchanged. By analogy, the rapid transitions we observed between variable and stereotyped song elements do not imply that the *range* of variability within each song element can be changed quickly, or reversed.

4.4.2 Hierarchical development

Taking articulatory dynamics into account, however, can complicate matters, since manipulating one vocal event could affect the acoustic states of neighboring events, and undo the learning. In this respect, learning temporary coarse song structure prior to the learning of temporary fine structure could potentially decrease such interactions and facilitate learning. As we have seen in Chapter 1, such transition from coarse segmentation of continuous action to fine segmentation may solve the problem of the trade-off between memory load (too fine segmentation) and the conflict between consolidation and exploration, which arises when segmentation is too coarse.

There is some evidence from previous studies that in articulation birds exposed to the normal song model (tutored birds) can learn more fine time structure than birds that have never heard a song model (untutored birds) (Méndez et al., 2010). Our results of exploratory variability analysis corroborate this view, since syllable types (clusters) became detectable in our data a few days prior to the appearance of robust intra-syllabic events. As we will see in the next chapter

(Chapter 5) untutored birds do not show any signs of hierarchical development of their song (although the song does undergo some development and eventually consolidation).

4.4.3 Future directions

The results presented in this chapter are largely based on correlations between variability and either global or local error. The main result shows that correlations with the local error are stronger, supporting the hypothesis that errors from the song template are computed locally to gate vocal exploration dynamically in very short time scales. It would be interesting to further study the causality between vocal changes and exploration. Would exploratory variability increase when new sub-syllabic song elements are added to syllables? We have tried tutoring birds with sequential training analogous to the AAA → ABAB, in which a new sub-syllabic element was added to the "target". However, we failed to induce intra syllabic changes in those birds. Possibly, the source syllable was too consolidated at the time when we altered the target syllable.. It is also possible that peripheral dynamics (e.g., co-articulation) imposed constraints were too strong to allow alteration of that syllable. However, this is not likely given the results by Tumer & Brainard (2007). A potentially better approach might be to set up an experiment analogous to Tumer and Brainard (Tumer and Brainard, 2007) where, as noted above, negative reinforcement was used to motivate birds to alter the fundamental frequency of a short song element. If such procedure is done in young birds, with plastic song, it is possible to test if the variability of that short song element will be higher than in the neighboring elements, when template match is not required. The alternative hypothesis is that the reason for our failure to induce intra syllable changes by altering the target is not due to motor constraints, but perhaps perceptual.

Chapter 5: Song learning at the articulatory level: the development of respiratory pressure patterns

Data used in this chapter were collected by Primoz Ravbar, Franz Goller and Brent Cooper at the Vicario lab & Goller lab. EMG data were collected by Franz Goller at the Goller lab. Pratik Shukla also contributed to data collection and provided major help with chronic recording maintenance. Xiao Jianqiang improved the recording techniques used in this chapter.

5.1 Background and Rationale

In previous chapters we explored the development of a syllable by analyzing its acoustic structure. We showed that birds can regulate variability across syllabic renditions such that parts of the song that need to change more (because they are further from the target) receive higher exploratory variability. The techniques were developed to track various parts of a syllable from their developmental origin onwards. We showed that the scope at which vocal exploration takes place could be quite narrow. We know (Kao et al., 2004) that perturbation of syllabic structure by the "learning pathway" – AFP – can be very precise, both in time and in the specificity of acoustic features affected.

But while we know that birds can manipulate song at short time intervals it remains less clear how the ability to do so develops. How does the *skill* to perform, modify and explore song elements develop? (We define *skill* as the ability to manipulate the song at short time scales.) It would be worthwhile to know, for example, which sets of muscles are involved in song production and how does the activity of those sets of muscle develop with song learning. Unfortunately analyzing the acoustic structure alone can tell us little about muscle activity.

This was the main motivation for identifying a physiological variable that can be followed developmentally to study basic processes that underline the emergence of motor skill. This variable is *respiratory pressure* and can be directly related to activity in a particular set of muscles - those controlling respiration (Suthers 1999; Goller 2002). In this chapter we will discuss the methods used to record and analyze the respiratory pressure patterns and see how these methods can be used to measure development of motor skill. But first let us examine the background of song production.

5.1.1 Major components required for song production

The song is produced by coordinated activity of the vocal apparatus, which include three major components:

1. *The Syrinx*
2. *The respiratory system*
3. *The beak and the upper vocal tract*

These components have to be well orchestrated during singing. Now we will present the functional role of each component.

The Syrinx

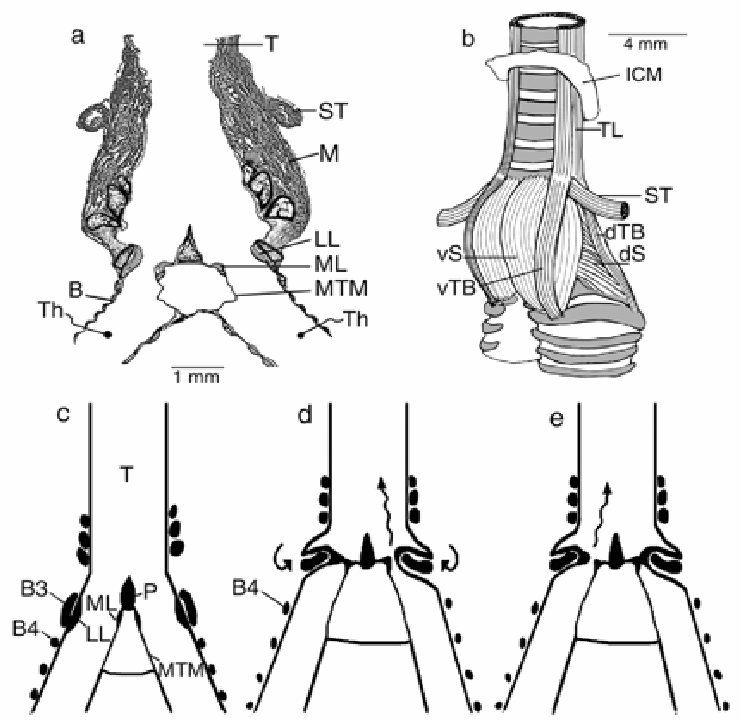

Figure 5.1 Schematics of the syrinx. See text.
Taken from Suthers et al., 1999

Syrinx, the vocal organ of songbirds, is located in the interclavicular air sac between the two bronchi and the trachea (in zebra finches). It is a complex organ composed modified cartilage and six muscles at each side. Those muscles are responsible for adductions (closing) and abductions (opening) of labia in each bronchus. The vibrations of the labia result in sound. Labia on both sides of the syrinx are fully abducted during silent respiration (Fig. 5.1*C*). During singing labia adduct (which increases the resistance in the bronchus and creates the right conditions for sound producing oscillations (Fig. 5.1*B*). Also, when labia are completly adducted they can act as vents, closing the ipsilateral bronchi (Fig. 5.1*E*). Extensive EMG recordings during singing revealed functions of different sets of muscles in the syrinx (Goller and Suthers, 1996). The *dorsal muscles* (dS and dTB, Fig B) are involved in closing (adduction) the bronchus, decreasing the air flow through the syrinx (Suthers et al, 1999). When these muscles contract, they adduct the labia by moving the cartilage of the bronchus. *Ventral muscles* (vS, vTB), on the other hand, control the frequency of sound my manipulating the tension of labia (Goller and Suthers, 1996).

Interestingly, the two sides of the syrinx can (in some species) produce sound quite independently from each other (Suthers, 1997). This "lateralization of song control" increases production capabilities of some species of songbirds: the two sides of the syrinx are often specialized for different frequency ranges. In domestic canary, for example, the right side produces high frequency parts of "FM sweeps" (these are syllables or parts of syllables where mean frequency is decreasing as to produce a "sweep" on a sonogram), and the left side produces the low frequency parts (Suthers et al, 2004). In other species, like in brown headed cowbirds, songs are produced by alternation between left and right side of the syrinx (Suthers and Roderick, 1997). In the zebra finch, however, most song syllables are produced by co-operation between the two sides of the syrinx, except from high pitch notes which are produced by the right side only (Goller and Cooper, 2004). But this lateralization of song production might also pose additional challenges of coordination; now the motor system has to coordinate not only respiratory muscle activity with the syrinx but also the two halves of the syrinx with each other.

The respiratory system

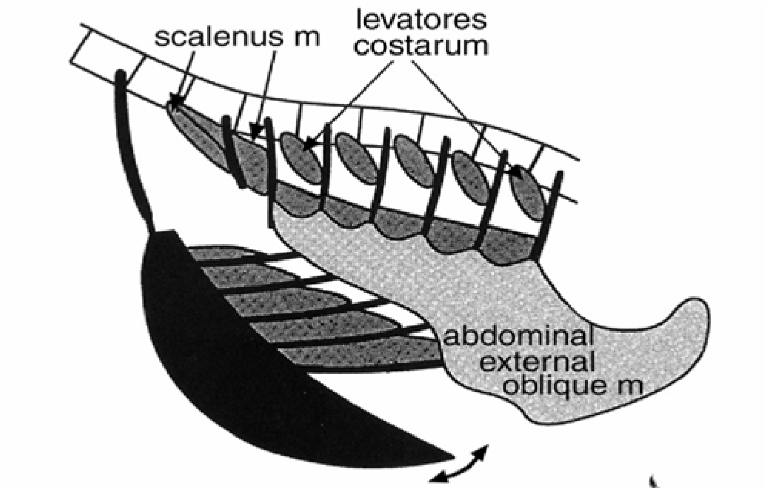

Figure 5. 2 Abdominal muscles are involved in production of expiratory pressure. The production of inspiratory pressure is controlled by *scalenus m*, *levatores costarum* muscles, while expiratory pressure is produced by abdominal muscles. *Taken from Suthers et al., 1999*

The respiratory pressure is produced by a set of about a dozen muscles acting on a complex of interconnected air sacs. The activity of these muscles has to be precisely coordinated with the syringeal muscle activity. In contrast to syrinx, the lateralization of song control does not apply to the respiratory pressure (Goller and Suthers, 1999). As such, the pressure amplitude is a global articulatory state, common to both bronchi. In birds, both inspiration and expiration are active processes (Konishi, 1965). The expiration pressure is produced by the activity of abdominal muscles, while the inspiratory pressure is controlled by a set of thoracic muscles (*scalenus m*, *levatores costarum*, see Fig, 5.2 above) (Suthers et al., 1999).

How does the respiratory pressure relate to the vocal output? When the air sac pressure is measured during singing we can observe a number of interesting properties. Unlike in human speech the respiratory pressure amplitude follows closely the acoustic structure of the song. The pressure signal is composed of repeated pulses of expiration which we will call "EPs" (expiratory pulses). EPs are correlated with the acoustic signal (Fig 5. 4). On a coarse scale the pressure correlates with syllable boundaries, but on a finer scale, it correlates with intra-syllablic structure. Note that pressure shape reflects, to a large extent, the input *to* the syrinx and not a combined effect of respiratory muscles and syringeal resistance. That is, EP shape is not significantly determined by changes in syrinx resistence (Cooper and Goller, 2004). This is shown in Figure 5. 3. Fine structure of the EPs is not affected by fixing of the syrinx such that it stays always open. Under this condition the resistance of the syrinx does not change and can therefore not contribute to the finer structure of the EPs observed. Notice that "singing" is silent under those conditions, while the pressure signal is not affected (Fig. 5. 3). This is an indication that EP shape is determined by the respiratory system alone. This fact is critical for the main aim of this chapter, namely to relate the development of the skill (ability to manipulate song at short time-scales) to the activity of muscles.

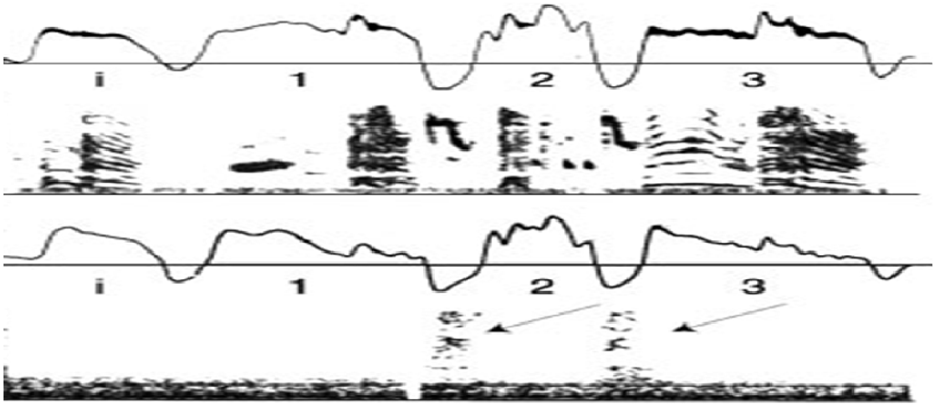

Figure 5. 3 The fine temporal structure of respiratory pressure does not result from the activity of the syrinx. The top trace is showing respiratory pressure pattern during vocalization (sonogram bellow). After muting procedure where the air resistance was set constant the pressure pattern retained all of the fine temporal structure (lower pressure trace; the sonogram after muting bellow). *Taken from Cooper and Goller, 2004*

This highly structured pattern of respiratory activity must in turn be tightly coordinated with syringeal activity, ensuring that the syrinx receives just the right amount of pressure for producing the correct sub-syllabic acoustic structure.

Another interesting feature of respiratory pressure is that while most animals exclusively vocalize during positive respiratory pressure phase (e.g. during expiration), birds also (albeit not often) vocalize song syllables during inspiration. Acoustic notes produced during inspiration are observed more often in early, non-crystallized songs (Goller and Daley, 2001). The song syllables are separated by short, usually silent periods during which the bird is inhaling. These short inhalations are called “minibreaths” (Calder, 1970) and are much deeper than inhalations produced during quiet respiration.

These properties make pressure especially important for the temporal organization of the song. If a singing bird is interrupted by a strobe light, the song usually (but not always) gets truncated usually *between*, but sometimes also *within* a syllable (Cynx, 1990). These “continuation to completion” effect is even stronger at to the level of expiratory pulses (EPs) observed on air sac pressure (Franz and Goller, 2002). Thus, when the song is interrupted by the strobe light, it breaks down *only* between the cycles of inspiration and expiration. In a sense the respiratory pressure determines the temporal structure of song and the motor units of its production it could be referred the *articulatory* component of singing.

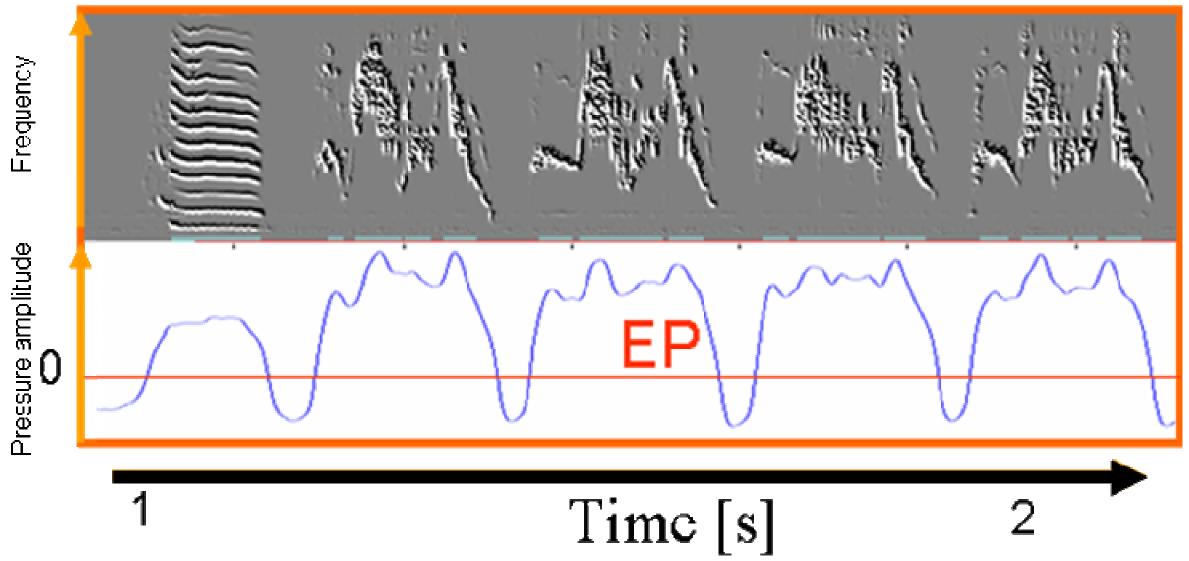

Figure 5. 4 Pressure amplitude follows closely the acoustic output of song. Note that pressure follows the sub-syllabic structure of acoustic syllables (repeated units in the spectrogram above).

The beak and the upper vocal tract

Another motor system, and a major component of song production, that has to be integrated with the syrinx and respiratory systems consists of the beak and the upper tract. It was shown that beak movements during the song are stereotyped (Goller et al., 2004). In zebra finches the beak gap is also well correlated with the peak frequency of different notes in a syllable, namely large beak gape usually results in removal of high frequencies from harmonic stacks. This may indicate that opening the beak shortens the upper vocal tract, thereby changing its filtering properties. Beak gap is also closely correlated with the amplitude of respiratory pressure. This coordination between respiratory system and mandibular motor control is not well understood. A lot more is known about coordination between respiratory system and the syrinx.

5.1.2 How are respiratory muscles and syrinx muscles coordinated?

This is a subject of current research and we will summarize just one possible mechanism by which the nervous system is coordinating respiratory activity with the activity of syringeal muscles.

The forebrain nucleus RA (*robustus archistriatralis*) sends projections to pre-motor nuclei of both parts of the vocal apparatus: syringeal and respiratory. On the syringeal part of the vocal apparatus, neurons from ventral RA ipsilateraly project onto motor neurons in the hypoglossal nucleus (nXIIts). These motor neurons drive the activity of syringeal muscles, both ventral and dorsal. On the respiratory part of the vocal apparatus, dorsal RA projects onto RAm (nucleus retroambigualis) and RVL (ventrolateral medulla, controlling quiet respiration). RAm, in turn projects to both nXIIts (bilaterally) and onto respiratory muscles (Wild et al, 2000). More recent findings (Kubke et al., 2009) suggest that RAm consists of two distinct neuronal populations both projecting onto nXIIts, as well as populations projecting onto motor neurons of respiratory muscles. These various populations of RAm neurons are likely involved in coordination of the syringeal muscle activity with the activity of respiratory muscles, exciting and inhibiting the nXIIts during different parts of respiratory cycle. It is thought that RAm may play a role in gating of the nXIIts activity so that appropriate *phase delays* are established between syringeal and expiratory muscles. These delays allow for enough expiratory pressure to be created (by expiratory muscles) before syringeal vocalization starts (Suthers and Margoliash, 2002; Kubke et al., 2009).

Now when we presented all the major components involved in song production let us go back to our main aim in quantifying the development of skill. But first let us present the two conditions under which we will measure skill development.

5.1.3 Acoustic isolates vs. tutored birds

In order to quantify the skill development we compared birds that were tutored with a song model to those that were raised in acoustic isolation. The latter birds can hear themselves vocalize but have no memorized tutor song to copy. The rational for this choice of experimental groups is that we would expect more skill development in birds who attempt to imitate the song model, than in acoustic isolates. That is, we would expect tutored birds to develop in the direction of being able to manipulate their song at shorter time scales. Interestingly, however, song undergoes some developmental changes even in these birds, although their song has atypical features and is not completely stereotyped (Marler, 1985).

5.2 Methods and definitions

5.2.1 Recording of respiratory pressure during singing

In all the birds, we used a surgically implanted scilastic tube (a cannula) to record respiratory pressure continuously for periods from a few days to a few weeks, together with the acoustic signal. Under these conditions, birds sing spontaneously.

5.2.2 Analysis of respiratory pressure patterns

From these data, we identified pulses of expiration on the respiratory pressure signal. These pulses usually correspond to syllables in the acoustics (Franz and Goller, 2002). There are also inspiratory pulses (negative pressure), usually associated with silences separating the syllables. Therefore when we study the activity of respiratory muscles, resulting in expiratory pressure pulses (EPs), we can infer that this activity is in fact involved in syllabic development. And, not surprisingly perhaps, as the syllables develop so do their corresponding EPs. Thus the basic units of our pressure pattern analysis are the EPs.

We then quantified the developmental changes in EP waveforms (raw data) by using the Fast Fourier Transformation (FFT), to compute a power spectrum for each EP analyzed. This method will be described in detail in this chapter, but its essential use is to quantify the addition of increasingly fine temporal structure to the pre-existing coarse structure of the EPs, over the period of song development. This process of addition of fine structure to the coarse structure we call *pressure refinement,* which is our measure of motor skill. Most importantly, our method allows us to compare pressure refinement (motor skill acquisition) between juvenile and adult birds as well as between tutored and untutored birds

We found that, although the song of untutored changes, it does not show the same refinement as in tutored controls. Thus, refinement (addition of fine temporal structure to EPs) does not occur merely as a function of vocal practice, but appears to require guidance by an acquired "template".

5.2.3 Data collection

Once the singing activity was stable we started recording the respiratory pressure from a plastic cannula inserted into the thoracic air sac. Surgery was performed under gas anesthesia (isoflurane), the skin was punched just under the rib cage and a cannula was fixed into position by surgical suture and tissue adhesive. This procedure is relatively non-invasive and the bird resumes singing 1-2 days after the surgery. The cannula was then led out to a pressure transducer incorporated in a backpack attached to the back of a bird. Pressure data was digitized together with the sound using NIDAQ card and the Sound Analyisis Pro software (Tchernichovski et al., 2000) for training and recording. The pressure signal was acquired at the sampling rate of 44.1 kHz and amplified. Filters were set to record signals from DC – 3 kHz. We monitored the pressure signal during singing and quiet respiration. If the signal to noise ratio (SNR) decreased drastically we repeated the cannula implant surgery. Only the data with consistently high SNR were analyzed. This procedure allowed us to record respiratory pressure in individual birds for continuous periods lasting 4-14 days. Pressure data was analyzed using Matlab 7. Raw signal was averaged to millisecond resolution. Data with no singing activity or low signal-to-noise ratio were discarded.

5.2.4 Segmentation

The pressure signal was segmented by a stationary threshold set at zero pressure (the average of silent respiration). This way EPs (expiratory pulses) crossing the threshold are isolated. We used mean amplitude and duration to segregate EPs and cluster them into different types. Only EPs belonging to the same type (cluster) were used for analysis.

5.2.5 Alignment and averaging

EPs belonging to the same type were aligned by the beginning (the first point at which they crossed the stationary threshold). These EP segments where then averaged across renditions to produce the averaged shapes of EPs presented in Figure 5.11 of the Results. We tried alternative alignment methods but did not observe a significant difference in averaged shapes of EPs.

5.2.6 Fast Fourier Transformation (FFT)

Any complex signal can be decomposed into a number of sine waves of different frequencies. After EPs of the same type were aligned we high-pass filtered them, using differential filter. The input signal was padded with zeros on both ends as necessary so that power spectra of the EPs could be computed on the filtered data (Matlab 7 *fft* function). Averages of 100 spectra of sequential EP renditions were taken in order to remove non-repeating frequencies. These averaged spectra were then used to calculate their mean frequencies (our measure of motor skill): of the center of gravity of Fourier coefficients (power) was taken as an estimate of mean frequency.

5.3 Results

5.3.1 Development of EP structure during song learning

The developmental changes in a representative zebra finch song (in a bird tutored with a song model) that lead to a stable, stereotyped “crystallized” song are illustrated in Figure 5.5. This figure shows an introductory note and two different syllables that make up a song bout. As can be seen from the sonograms, the syllables become progressively more similar to the model as the song develops from Day 55 of age to Day 72 (training started on Day 43). These changes are accompanied by changes in the pressure signal (blue trace), as well as the acoustic amplitude signal (orange trace). The horizontal black line indicates zero respiratory pressure (room pressure or averaged pressure of quit respiration). This line is used as a threshold to segment the pressure trace into positive pressure pulses (pressure above the zero-line) and negative pressure pulses (below the threshold line). We refer to the positive expiratory pressure pulses as EPs. Nearly all vocalization is produced during the EPs although a subclass of syllables can be produced on inspiration (Goller et al., 2001). Fine temporal modulations can be seen not only on the acoustic amplitude but also on the respiratory pressure, suggesting the contribution of respiratory as well as syringeal control to song production (see Discussion of this chapter).

Importantly, the modulation of the respiratory pressure signal increases over the period of development from Day 55 to Day 72, as is evident from the pressure signal (blue trace) in Figure 5.5*A*. The changes are seen in the superimposed EP segments in 5.5*B*. The red arrows in 5.5A indicate the points where new valleys and peaks have appeared.

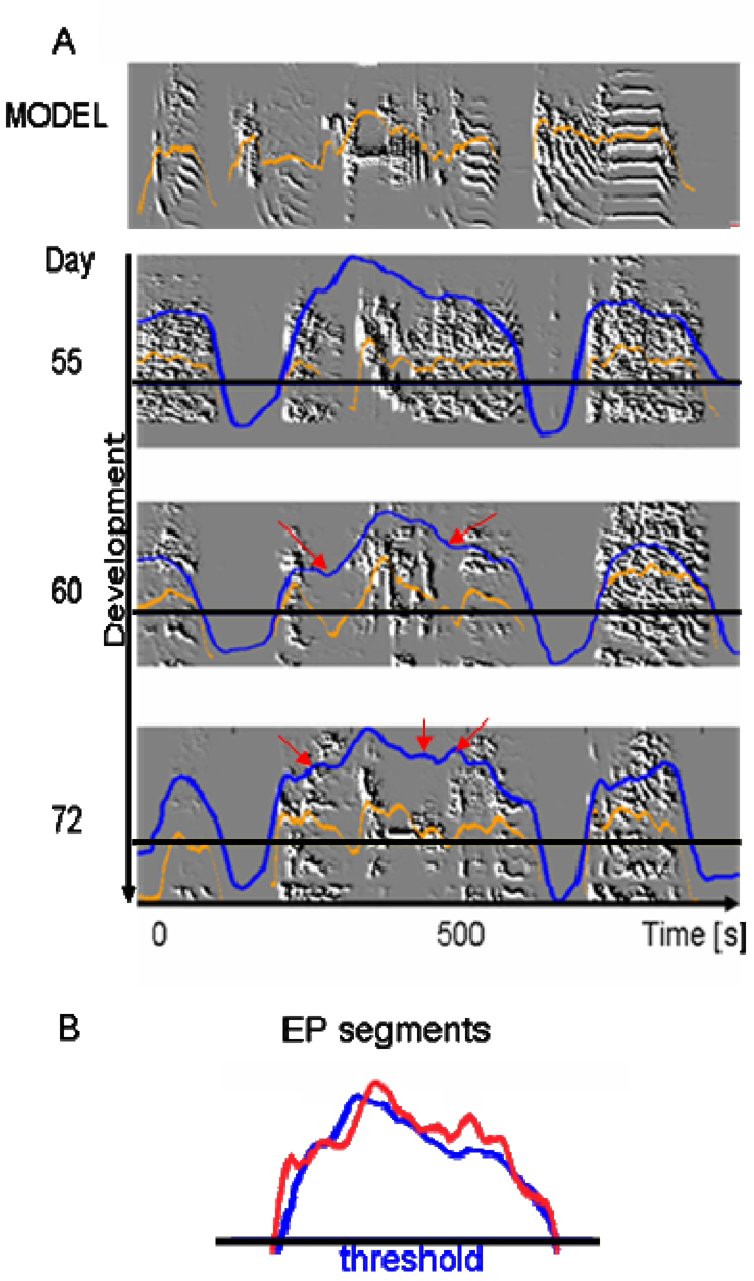

Figure 5. 5 Song development. ***A,*** sonograms (vertical axis is frequency; horizontal axis time) show song development from day 55 to day 72 after hatching. The blue trace shows the respiratory pressure and the orange trace represents sound amplitude. The horizontal black lines identify zero pressure also used as a threshold to segment expiratory pulses (EPs). The duration of these song bouts is 1.0 sec. Red arrows point to examples of added fine structure. ***B***, EP segments from Day 55 (blue) and Day 72 (red) are superimposed for comparison.

Changes in EP waveform over the recording period were seen in all birds, but differed dramatically between tutored and acoustically isolated birds. Figure 5.11 compares early and late examples of vocalizations from three tutored and three untutored birds (Fig. 5.11*A* and 5.11*B* respectively). The averaged EP shapes (across 100 renditions) are shown for each bird (blue and red traces represent early and late recordings, respectively), next to their corresponding sonograms. For each bird, the upper sonogram (in blue frame) is a sample taken from the beginning of a pressure recording period and the lower sonogram (in red frame) is an example obtained from the end of the recording period. For each bird we also present the EP spectra (produced by FFT analysis of EPs), which we use to quantify the amount of temporally fine structure present in the EPs. It is now time to describe this method as is done in the following section.

5.3.2 Application of Fourier Transformation to quantify the presence of fine temporal structure in the EPs

As noted earlier, when EP structure develops over the period of song learning, temporally finer structure is added to the pre-existing coarse structure. We refer to this process as pressure refinement or motor skill acquisition. The addition of fine structure can be quantified by first decomposing the EPs by Fourier Transformation. The spectra thus obtained can be collapsed across renditions. The averaging of EP spectra is done in order to determine the repetitive frequencies e.g. the fine structure of EPs that is consistently repeated across renditions (see Methods and definitions for details).

We can demonstrate this method by a simple simulation. Any signal can be decomposed in its frequency components by Fourier Transformation (FT). Take a simple sine wave with a frequency of 10 Hz and decompose it by FT as shown in Fig. 5.6 below:

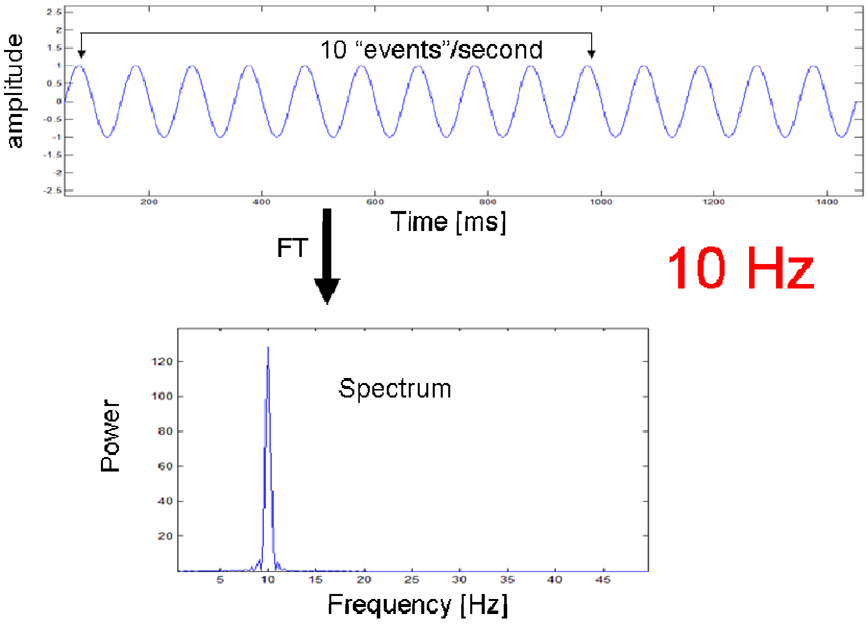

Figure 5. 6 Fourier transformation of a 10 Hz signal. See text.

The obtained spectrum has one peak at 10 Hz. Now let us look at another sine wave of 1 Hz frequency and decompose it (Fig. 5.7):

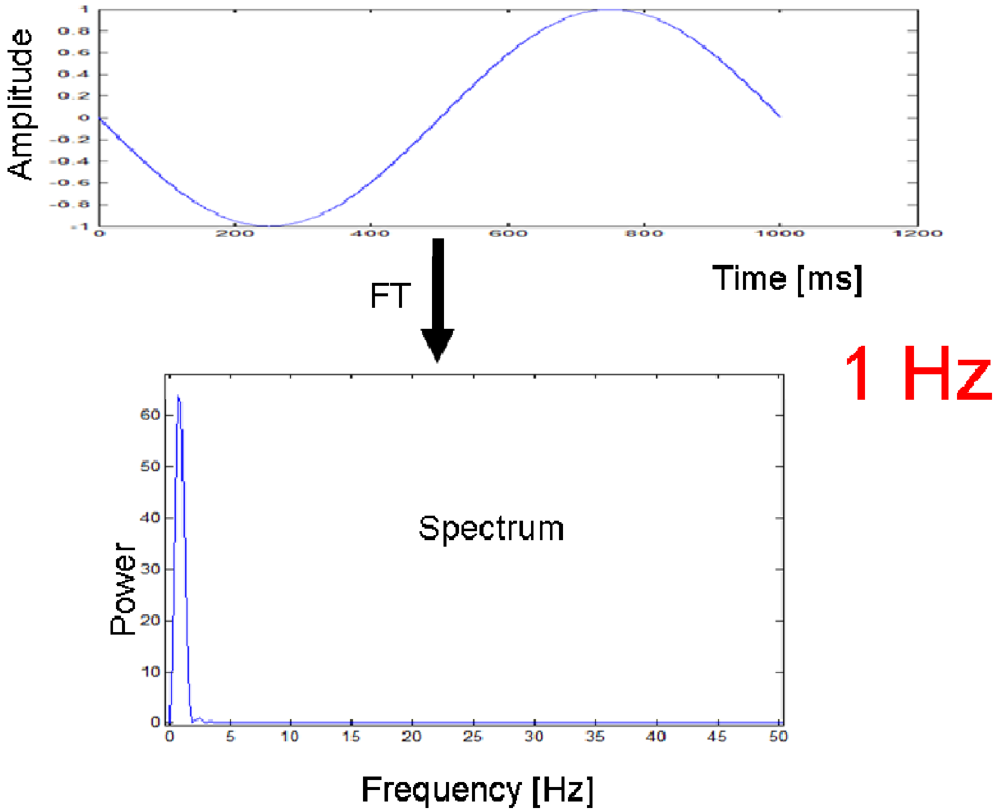

Figure 5. 7 Fourier transformation of a 1 Hz signal.   See text.

This signal has a single peak at 1 Hz. Next let us *add* the 10 Hz signal to the 1 Hz signal to obtain the *composite* signal (Fig. 5.8):

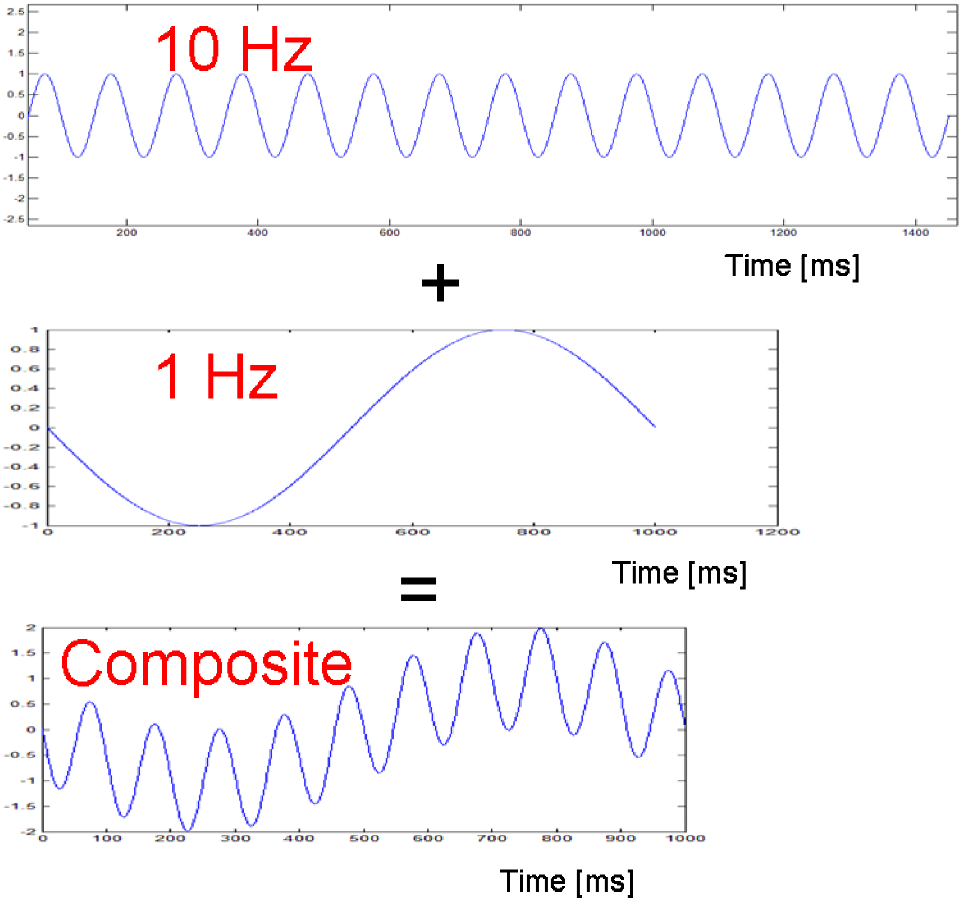

Figure 5. 8 Adding signals with different frequencies results in a composite signal. See text.

This composite signal can then be decomposed by FT to obtain its spectrum as shown below (Fig 5.9):

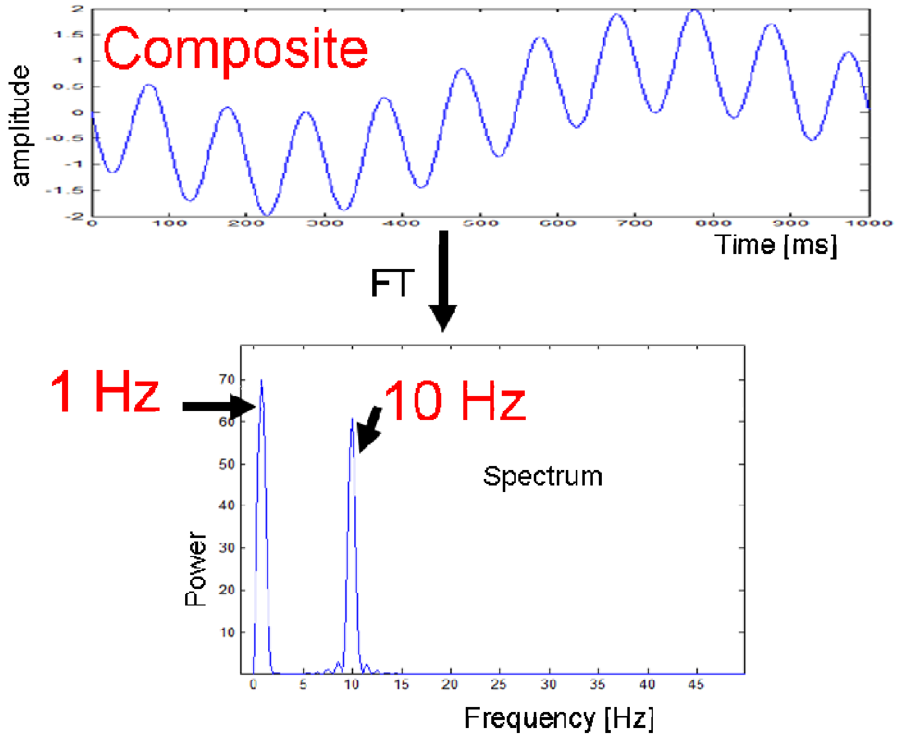

Figure 5. 9 Fourier transformation of a composite signal. See text.

The spectrum of the composite signal has two peaks, one for each component: 1 Hz and 10 Hz peaks.

We can represent the coarse structure of EPs with the low frequency signal (low level of structure modulation) and the “refinement” of EP structure with the *addition* of a high frequency signal to the pre-existing coarse structure. Therefore the refinement of structure should result in a new, higher frequency peak in the spectrum as summarized in Fig 5.10 below:

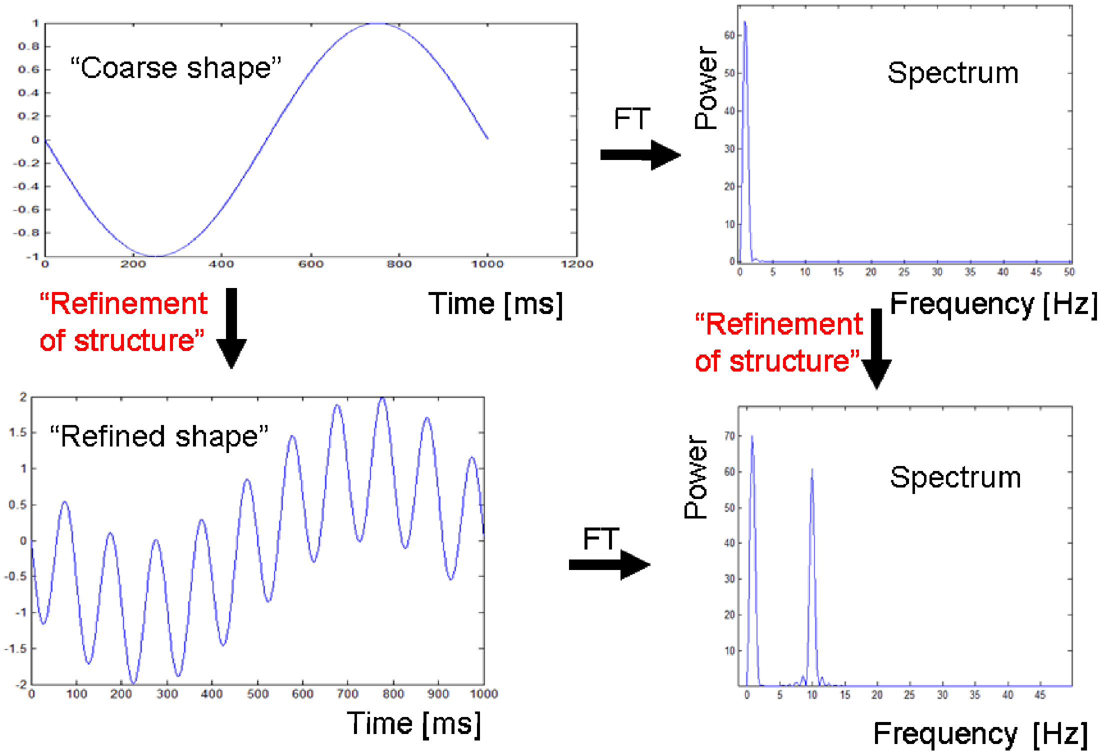

Figure 5. 10 Adding a higher frequency "fine" structure to the low frequency "coarse" structure results in a shift of spectral power towards higher frequency. See text.

Thus we predicted that with refinement of EP structure of trained birds the spectral power would shift toward higher frequency. On the other hand, spectra of isolated birds, where no or little structural refinement happen should show no such shift of spectral power. (To eliminate the effect of noise that could potentially get added to the EP measurements we calculated means across spectra so that only the repeated frequencies "survive".)

5.3.3 Structural refinement happens predominantly in tutored birds

In Figure 5.11 compare the spectra of isolated birds to the spectra of tutored birds. Note that even at the end of the recording period (red spectra) untutored birds have less power concentrated in lower frequency ranges, which indicates that their EPs do not have as much fine temporal structure as the EPs of tutored birds. Moreover, isolated birds do not show *improvement* of motor skill, as measured by the shift of power of EP spectra towards higher frequency ranges, while trained birds do show a shift, indicative of pressure refinement.

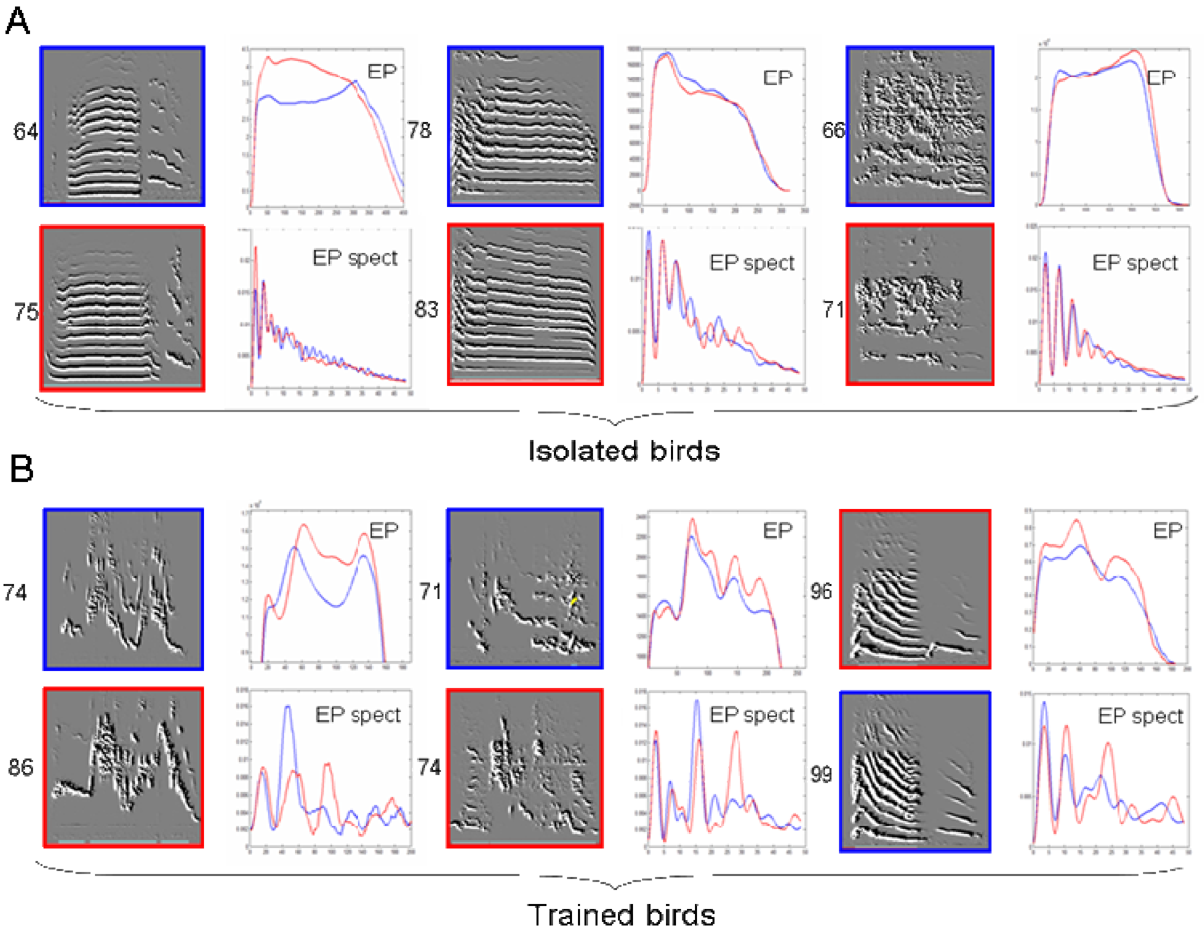

Figure 5. 11 Comparing early to late EPs and their spectra in trained birds and isolates. ***A,*** Three examples of isolated birds are shown. Blue traces indicate early samples, red traces late samples. The corresponding sonograms are shown (blue frames indicate early samples, red frames late samples. ***B,*** In trained birds spectral power shifts from lower frequencies in early recordings (blue curves) toward higher frequency ranges in late recordings (red curves).

To quantify these differences further we performed this spectral analysis across five tutored birds and compared them to 5 un-tutored birds. Figure 5.12 shows the mean frequency of spectra for tutored and untutored birds. Empty circles indicate beginnings and ends of recording periods for individual birds. Each mean frequency is calculated as an average of mean frequencies of all types of EPs analyzed per individual bird (typically 1-2 types of EPs). Red circles and links indicate 5 tutored birds, while blue circles and links represent acoustically isolated birds. It is possible that the difference of EP structure refinement that we measure could be explained with the particular song model we used to tutor the birds with. So we also analyzed end-point samples of EPs from birds tutored with different song models (filled red circles). We found that the mean frequency of all 10 tutored birds was higher than the mean frequency of all isolated birds (Fig. 5. 12*A*).

While isolated birds may show less fine structure of their EPs than tutored birds, they might still be able to improve this measurement during a period of development. However, we found that across birds this mean frequency shift, indicative of refinement process, was significant only in tutored birds (ANOVA1; p=0.0378) but no significant change (ANOVA1; p=0.6457) was observed in isolated birds (Fig. 5. 12*C*). Thus we detected no improvement of EP refinement in untutored birds.

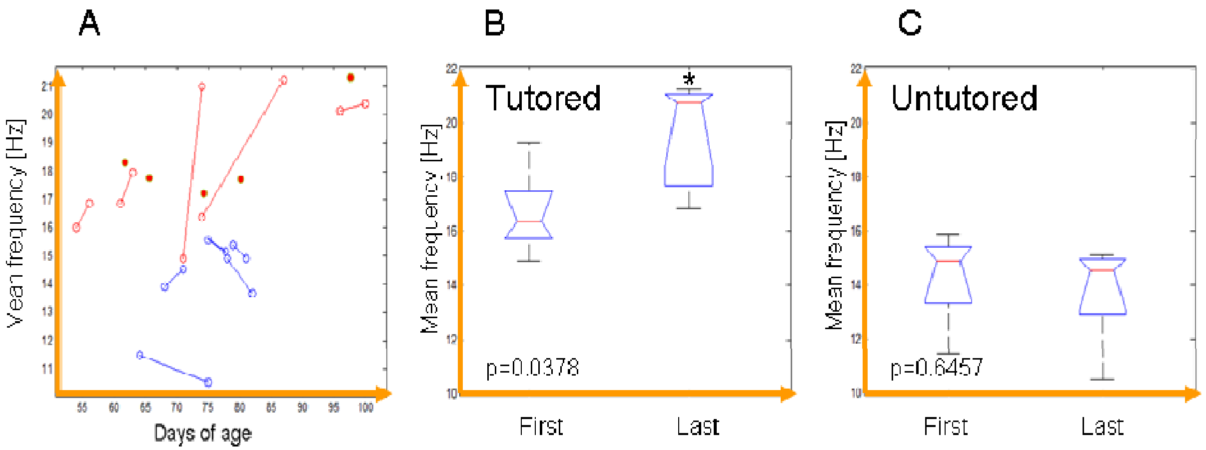

Figure 5.12 The mean frequency of EP spectra increases in trained birds and is overall higher than in acoustically isolated birds. ***A***, Circles indicate mean frequencies at the beginnings and ends of recordings (see methods). Red: trained birds; blue: isolates. Full red circles: samples of birds trained with different models. Mean frequency shifts across trained birds and is overall higher ***B***, than in isolates ***C***.

We wanted to characterize in what frequency ranges addition of fine structure happens, which corresponds to the *time scales* at which a bird refines his song. Figure 5.13*A* compares spectra from the beginning and end of recording in one tutored bird, and this example shows how power is added to certain higher frequency ranges (red fill), while power is lost in lower ranges (blue fill). Figure 5.13*B* shows the changes in power spectra in 5 tutored birds (different colors of circles represent the 5 subjects). Data points above the zero line indicate an increase of power for the corresponding frequency range. The points below the zero line indicate decreases of power. Figure 5.13*C* shows a box-plot of the data from Fig. 5.13*B*. Power significantly increased in high frequency ranges of 20-30 Hz and 40-50 Hz (*t-test*; *p-value* = 0.0244 and 0.0143 respectively), with the maximal increase in 20-30 Hz range. The addition of fine structure in these frequency ranges is also significantly higher than in 1-10 and 10-20 Hz ranges.

In terms of time-scales this result means that the greatest addition of fine structure to EPs spans from ~30-50 ms in length. For example, this would correspond to the broader three peaks of the middle EP in Figure 5.5*A* (note the deepening of the valleys – red arrows, separating these 3 parts of the EP, from Day 55 to Day 60 in Figure 5.5). The addition of fine structure was also significant (*t-test*; *p-value* = 0.0143) at the shorter time-scales, from 20-25 ms (40-50Hz in Fig. 5.8*C*). This finer structure could correspond to the peaks and valleys such as those added to the EP on Day 72 in Figure 5.5*A*.

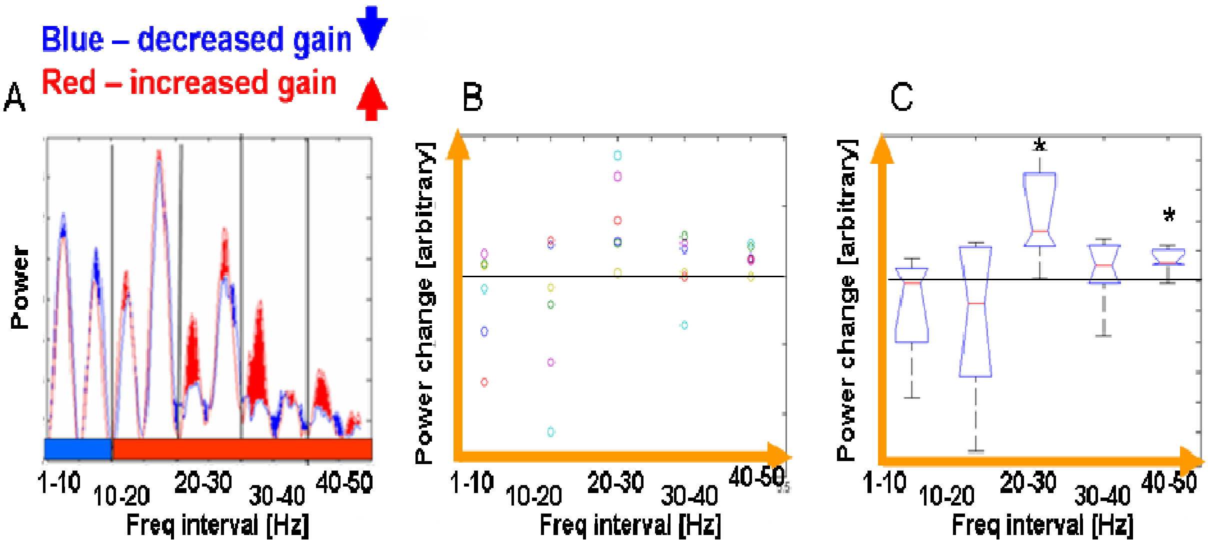

Figure 5.13 Identifying the frequency intervals where power changes occur in tutored birds**.** ***A***, Example of power changes in one bird (red: increase; blue: decrease). ***B*****,** Power changes in 5 EPs of trained birds in 10 Hz intervals. Units of change are arbitrary. ***C*****,** Box plots of power changes across 5 EPs of trained birds. Significant changes of power are indicated by "*". The highest increase of power was in 20-30 Hz interval, corresponding to time scale of 30 – 50 ms. We also see significant increase of power in 40-50 Hz interval indicating the addition of temporal fine structure as short as 20-30 ms.

5.3.4 Changes in respiratory activity are closely associated with changes in syringeal muscle activity

As already noted, previous studies have shown that fine temporal structure of respiratory pressure is not affected by the resistance of the syrinx (Cooper and Goller, 2004). This fact is critical for the interpretation of our results because we wanted to associate the structural refinement with the activity of a particular set of muscles (respiratory muscles). We also wanted to know weather the refinement of respiratory pressure would precede the refinement of activity other components of the song production system (see Discussion).

To find the answer we recorded the activity of the syringeal muscles by EMG, over a period of song development, in three birds. We cross-correlated the developmental time courses of parts of EPs with the same time courses in the EMG signal where developmental changes were observed and found no lag (median across three birds: r=0.47, lag=0). Figure 5.14 shows the aligned EPs (first column), acoustic features and the EMG in two birds. Note that especially where the correlations between EPs and EMG are strong (lower row) there is no apparent lag between the two correlated signals signals. Although this is a small sample of birds, the correlation between changes in EP shape and changes in EMG signal from the syrinx is indicating that structural refinement is probably coordinated between the two components of song production.

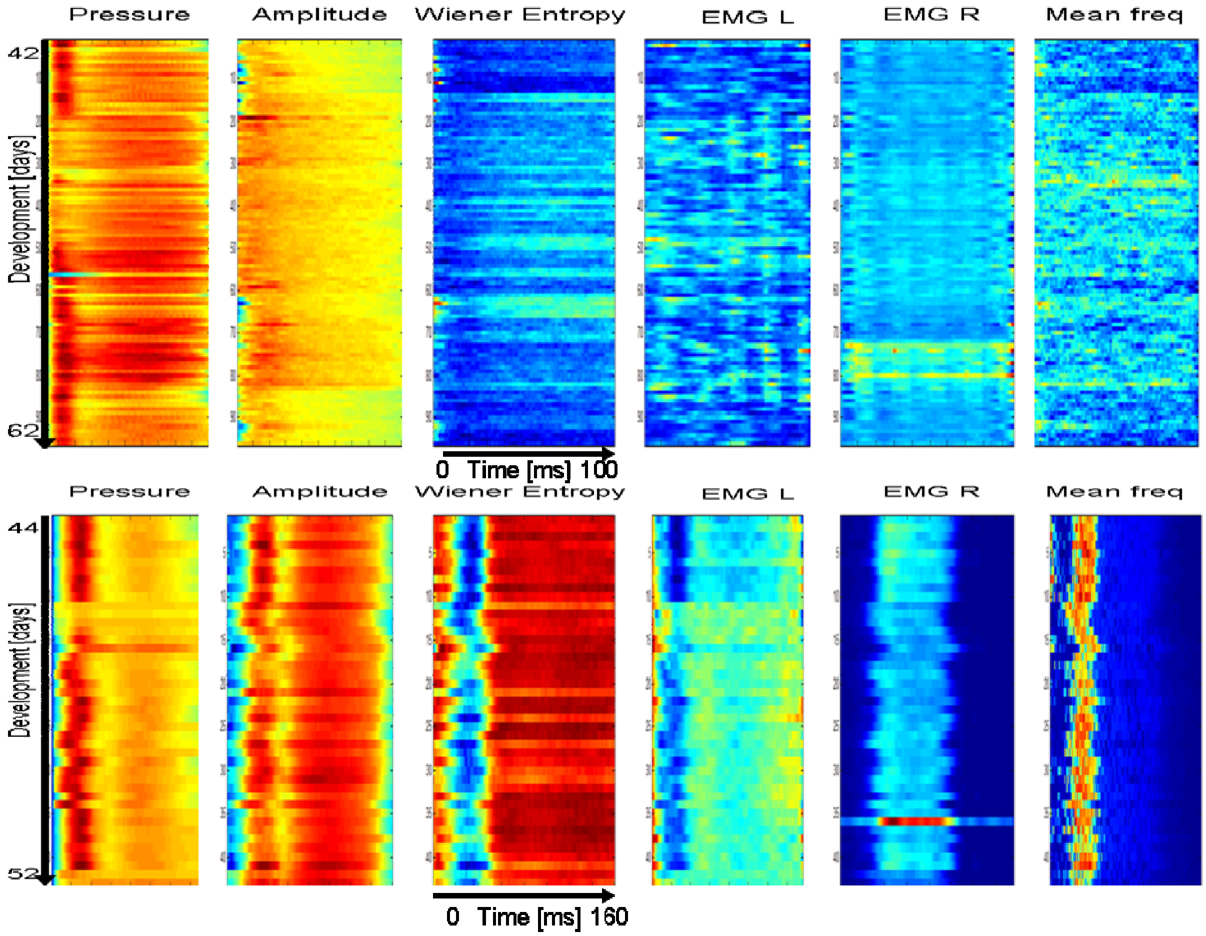

Figure 5.14 EMG activity is correlated with respiratory pressure with no lag. Respiratory pressure and EMG was recorded together with sound (in the laboratory of Franz Goller, University of Utah). EMG was recorded from both left and right muscles of the syrinx (EMG L and EMG R). Three birds were analyzed (two are shown). In cases where respiratory pressure and EMG signals were strongly correlated (lower row) the lag of cross-correlation was zero (compare pressure to EMG L in the lower row).

5.4 Conclusion and discussion

In this chapter we compared the syllabic development between tutored and untutored birds. Here are three of our main findings:

1) Tutored birds add more fine temporal structure to their EPs (expiratory pulses) than do untutored birds, over the course of song development. This confirms the previous finding (Méndez et al., 2010).

2) There is no indication that untutored birds add any fine temporal structure to their EPs during song development, although changes in their EP shapes do occur. These developmental changes do not seem to include addition of new peaks and valleys or other increments of modulation (such as peaks getting higher, valleys getting deeper), that could be detected by our method.

3) In the samples of EMG data from syringeal muscles where developmental changes of EMG shapes were closely associated with changes in the EP shapes, we did not find and lag of either signal. While due to small sample size (three birds) this result is not fully conclusive it could suggest that the addition of fine structure has to happen on both, respiratory and syrengeal level of song production system simultaneously.

In human speech respiration has to be coordinated with articulation (Conrad and Schönle, 1979) and the control of the breathing pattern is a critical art that has to be learned by singers (Sundberg Johan, 1993). The right amount of air has to pass through the articulatory and vocal structures in order to produce the desired sounds. On the other hand, the gas exchange in the lungs has to be uninterrupted during speech or singing. However, in human vocal behavior there is no need for fine temporal coordination of respiration. Instead, most of the articulation is performed by articulatory structures controlled by different muscle systems. As the result human respiratory pattern lacks time complexity found in some songbirds.

In zebra finches, on the other hand, the respiratory activity during singing is itself involved in articulation and, as a result of this, possesses a great temporal complexity that corresponds to the acoustic complexity of the syllables (Wild et al, 1998). This is perhaps surprising because, in principle, the fine articulation of song structure could be carried out by the syrinx alone (this would be analogical to human articulation), which would pose a lesser challenge for precise temporal coordination between syrinx and respiratory muscles. If this was the case, then it might suffice for respiratory activity to determine the rhythm or the syntax of the song, while the fine intra-syllabic structure would be learned on the level of the syrinx. But in zebra finches temporal structure of respiration closely follows the complex acoustic structure of syllables. This would imply that the learning of fine temporal structure of respiration must thus co-occur with the learning of other motor gestures (syringeal activity and beak movements). We show that tutored birds add fine temporal structure to the coarse respiration pattern during the song development.

While the major advantage of our method is in that we can indirectly measure the activity of (respiratory) muscles rather than acoustic structure (which results from complex interaction between muscles, membranes in the syirinx and respiratory muscles), there is also a possible disadvantage. It lays in the fact that respiratory signal is smoother than some acoustic features (see Fig. 5.5) so we may be missing even finer time-scale segmentation of syllables than what we can see in respiratory pressure. To detect such fine time-scale events (and relate them to central activity) one could record chronically from nuclei such as RA or downstream from it. Our attempts to produce such data have not been successful.

5.4.1 Bias towards structure refinement in tutored birds

That the song of untutored birds is significantly different from “normal”, tutored birds has been known for some time (Marler and Sherman, 1985). The difference in the respiratory pressure pattern between tutored and untutored birds has also been established in a recent study (Méndez et al., 2010) by measuring the degree of EP modulation at the end-point of song development together with the syringeal muscle activation. But, because the study is focusing only at the end-point of development, it cannot establish weather the untutored birds also follow the trend of refinement of motor gestures over development, but to the lesser extent than tutored birds.

Our results suggest that while there is a measurable systematic bias in birds exposed to the song model toward addition of fine temporal structure to the respiratory pressure patterns, this bias was not detected in untutored birds. Thus the difference between the two groups is not merely in the degree of motor gesture refinement (which has been shown before) but in the *direction* of development (qualitative difference). The lack of motor refinement in untutored birds could be caused by the broad time scale at which the vocal exploration is carried out (and at which the song can be modified). Because there is no model to imitate, the “goal” of song development may not be locally defined at all. In terms of hierarchical learning, which we discussed earlier, isolates may be stuck at the lowest level, where only broad acoustic changes can be made. Indeed, our observations show that developmental changes of EP structure do occur in untutored birds (the song does develop) but these changes remain broad. In other words, the untutored birds may not be able to produce developmental changes at the short time scale.

However, as mentioned earlier, even the untutored birds sometimes produce songs with complex but variable temporal structure (e.g. their songs are inconsistent). One advantage of our method is that we detect only *consistently produced* fine temporal structure of the song (by first decomposing EP shapes by FT and then *averaging* the EP spectra, see Methods and definitions). Therefore any untutored birds with highly temporally complex but variable songs would still be detected as possessing low level of temporally fine structure.

While it is in principle possible that untutored birds can “catch up” with the tutored birds later in development, this is unlikely for two reasons: first, as discussed above, untutored birds do not show any bias in the direction of motor refinement, and the second reason is, that the acquisition of motor skill is thought to be limited by the “critical period”, when substantial developmental changes can still be made (Konishi, 1965). The oldest tutored bird that we recorded pressure from (the recording session took place between 94 to 99 days post-hatching) also shows the lowest amount of pressure structure refinement (Figure 5.12*A*). This would suggest that the end of the “critical period”, when refinement of motor gestures can still occur, was already quite close for this bird. Therefore, it does not seem that untutored individuals would ever “catch up” with the tutored ones.

The inability of untutored birds to “catch up” with the tutored ones is, however, probably limited to individuals but not communities. A recent study by Feher et al. (Feher et al., 2008) looked at the syllables of zebra finches that were raised in isolation from normal (“wildtype”) songs but not from each other, and demonstrated that (acoustically) parts of syllables (“vocal states”) tend to get shorter over several generations of “tutors” and “pupils”. In other words, there is a systematic bias of the whole population to culturally evolve the song with fine temporal structure.

One problem with identifying just the length of vocal states from acoustic signal alone is that this method will not likely detect all developmental changes. A potential advantage of our method is that we can detect any changes in modulation of EP structure, even if the lengths of vocal states do not change. Therefore any trend in untutored birds towards “normal” song should be detectable by our method.

Veit et al. (Veit et al, 2011) have recently shown that at least one aspect of respiratory structure in very young birds is established before the acoustic structure (namely the length of gaps between syllables is adjusted on respiratory pressure first). This study therefore indicates that in young birds, where AFP has the dominant control of song production, the learning is focused on the establishment of respiratory patterns. There is also anatomical evidence (Vicario, 1991) that there is some degree of segregation in nucleus RA between parts controlling respiratory system and parts controlling the syringeal muscles, making it possible that refinement of articulation would be learned separately from refinement of syringeal activity.

However, this body of evidence does not necessarily imply that on the level of the syrinx the acquisition of fine motor skill (such as learning of intra-syllabic events) lags behind the respiratory system. Even in the absence of acoustic evidence the refinement of activity of syringeal muscles may still be present. As discussed earlier in this chapter, the acoustic output is the final product of integration of respiratory, syringeal and beak activity. The imprecision of this integration may result in apparently delayed measurements of refinement on the acoustic level. Indeed, as our results from EMG recording of syringeal muscle activity seem to suggest, there is no apparent lack of synchronicity between developmental changes in the pressure pattern and changes in syringeal EMG activity. Therefore we suggest that learning of fine temporal structure of syllables occurs on all levels of song production at the same time.

Chapter 6: Collateral damage of vocal changes?

When you chop wood, chips fly Nikolai Yezhov (Joseph Stalin NKVD officer, justifying executions of innocents)

6.1 Rationale

In the previous chapter we saw how hierarchical learning might explain the evidence supporting transition from coarse to fine units of vocal change. We suggested that as the acoustic structure becomes more consolidated, the size of segments should decrease (higher granularity). Another possible advantage to decreasing of the segment size is decreasing co-articulation-like adverse effects during learning. As mentioned in Chapter 4 (Discussion) might arise when the learning of one part of a continuous action adversely affect the structure of a neighboring part. Here we present a few cases where such interactions across different parts of the syllable seem to take place during learning. The results presented here are preliminary and conclusions should be considered as hypotheses for further studies.

In order to test how vocal changes of one part of a syllable may affect other parts we developed some new methods for tracking vocal changes across development, which might be useful for future studies and are hence presented in details.

6.2 Methods

6.2.1 Tracking vocal changes by local realignment

In order to track the vocal changes in the syllable across development we have to align syllables, and then locally realign the intra-syllabic events so as to be able to observe how changes in one song elements might affect neighboring elements (Fig. 6.1). As in previous chapters we focus on Wiener entropy as the principal feature. Figure 6.1 is composed of aligned segments of a syllable, where color represents Wiener entropy levels. The shape of Wiener entropy values over the syllable changes across development (from Day 50 to Day 76), starting from three coarse peaks and two minima, into a much more complex shape. On the left the syllables are aligned by the first millisecond (real time). On the right side syllables are locally realigned using our method. Note that on the right side of Fig. 6.1 the x-axis is not time anymore. Instead the x-axis represents the warped time. This representation enables us to align each event of the syllable with the corresponding event in the previous syllable. Figure 6.2 demonstrates the method.

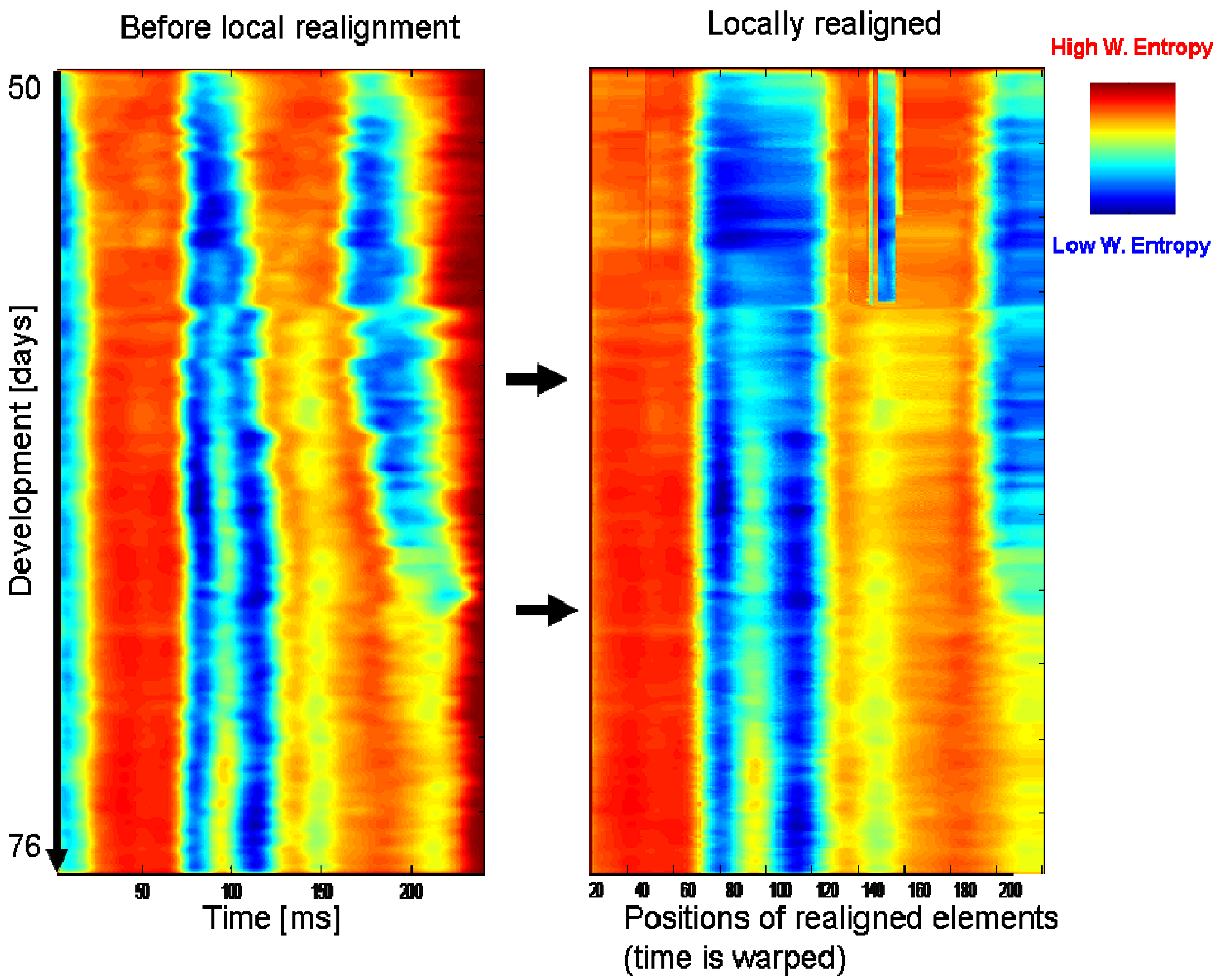

Figure 6.1 Results of local realignment. Wiener entropy is the feature used. The method aligns the syllables locally so that each event of the syllable is aligned with the corresponding event in a previous rendition of the syllables. See text.

We compute the maximum cross-correlation between two consecutive (in developmental time) segments of a syllable as the basis for realignment. However, not the entire syllables are cross-correlated but rather shorter segments (typically 30 ms long frames). The algorithm then "tracks" (according to best cross-correlation) each frame from the last day of development (in this case Day 76) to the first day of development. This is demonstrated in Figure 6.2.

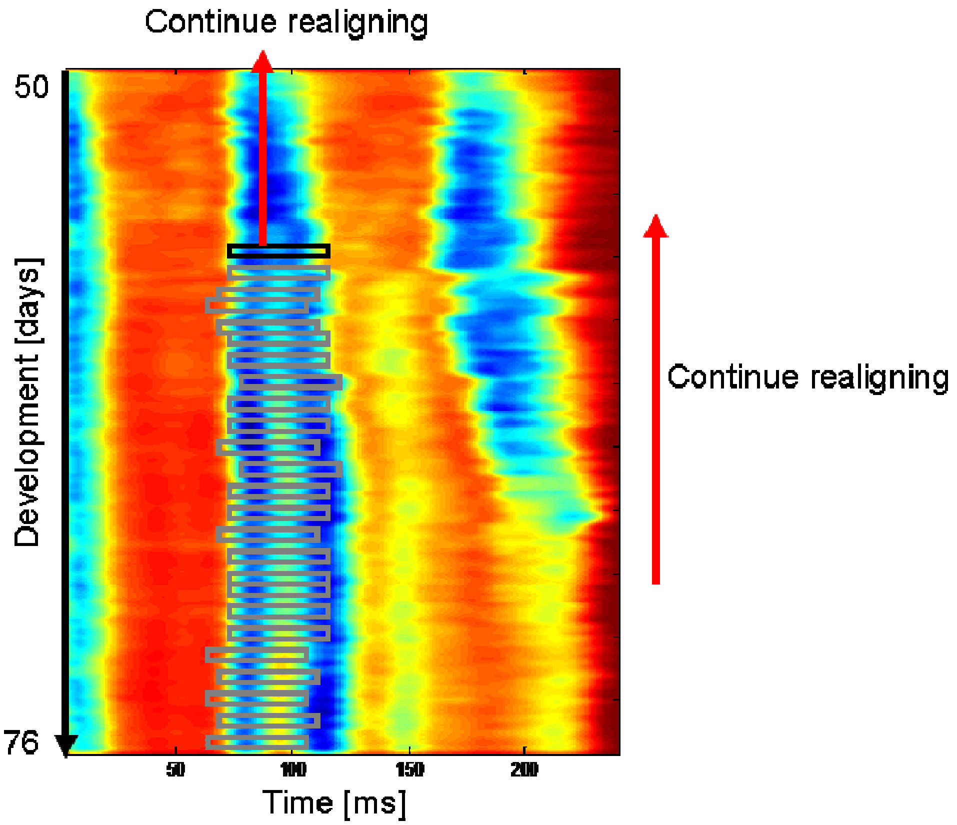

Figure 6.2 Algorithm for local realignment using best cross-correlation. Each frame (gray) is cross-correlated with the previous frame (in developmental time). The maximum cross-correlation is then used to realign the frames. Once the tracking algorithm reaches the beginning of development (Day 50 here) the entire process is repeated but the first frame (at Day 76) is moved for one millisecond in song time (x-axis). The final product of the algorithm is the realigned raster plot in Fig. 6.1 (right side).

The algorithm continues to locally realign frames from the end of development (Day 76) to the beginning of development. Once that point is reached the process repeats, but this time the first frame (at Day 76) is chosen one millisecond later in song time (the x-axis). The final product of the algorithm is the realigned raster plot shown in Fig. 6.1 (right side). Note in the realigned raster Figure 6.1 an artifact caused by the algorithm (right side, starting at Day 50 at about 130 on the x-axis). Such artifacts will be covered in the following figures.

6.2.2 Determining significant vocal changes in the syllable

We used locally realigned raster plots as the one shown in Figure 6.1 in order to measure the developmental vocal changes. For each realigned element (x-axis) we computed the amount of change over a period of one day, for every day of the development. Only the statistically significant changes were accounted for.

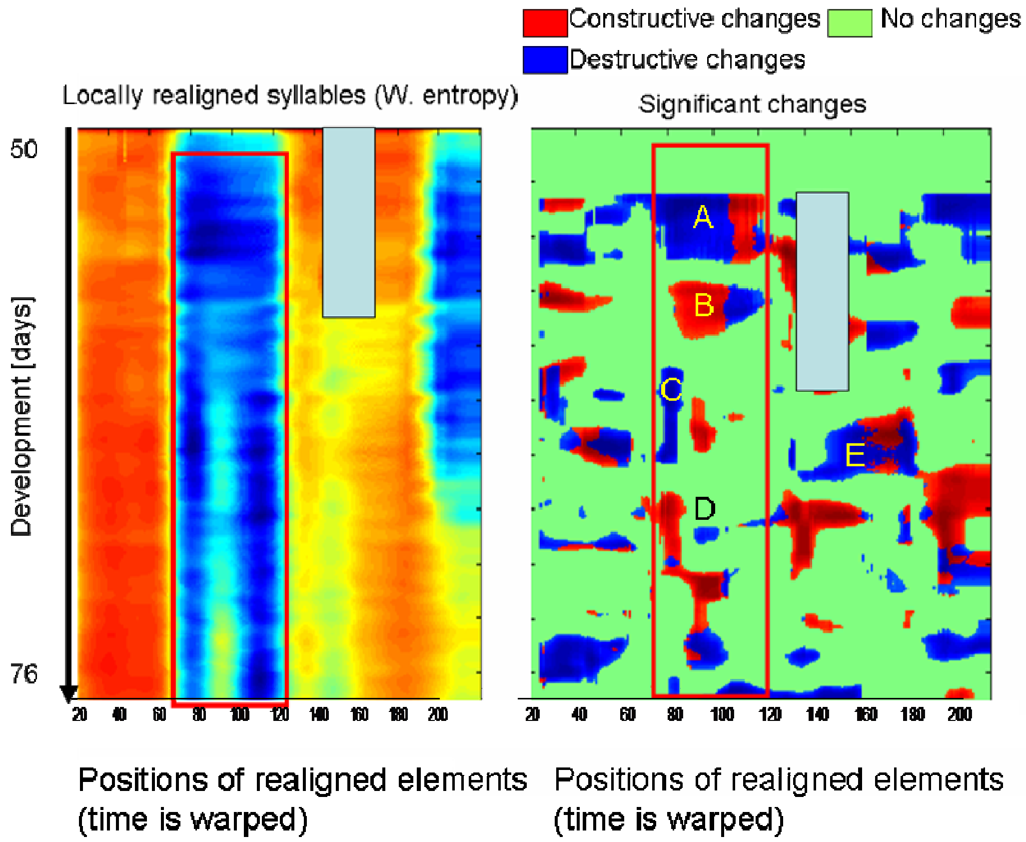

Figure 6.3 Constructive and destructive vocal changes. See text.

6.3 Results and Discussion

Figure 6.3 shows the realigned raster as produced by the local realignment algorithm described in the Methods (left side). The right side of Fig. 6.3 shows statistically significant vocal changes of realigned elements (all events in the syllable) as they occurred each day (one day was the frame in which changes were measured). Red colored changes represent “constructive” changes – where the syllabic structure becomes more similar to the song model that the bird is imitating. Blue areas in Fig. 6.3 (right raster) are significant vocal changes in the direction away from the song model (“destructive changes”). The green background of the raster represents the areas where no significant vocal changes happen (during each day).

Note in Fig. 6.3 that the changes are broad at first (at the beginning of development) but become narrower over developmental time. This is particularly noticeable in the area inside the red frame in the raster.

Compare the area in the red frame in the right raster in Fig. 6.2 to the same frame in the left raster plot. Notice in the left raster that a new event in the Wiener entropy was created. At first the framed area in the left raster is blue (low Wiener entropy) but later in development a peak of Wiener entropy appears. Now note in the right raster the broad vocal change that happens in this developmental period (labeled “A”). Most of this change is destructive (blue) but there is a short part of the change that is constructive (red part). Immediately after this (the next day) another broad change appears (labeled “B”) but now most of the change is constructive (red). In fact the change B seems to compliment the previous change A (and is also narrower). The change B happens when the peak in Wiener entropy appears (see left raster).

We speculate that early in development the bird makes broader changes as the segmentation is broad (see Chapter 5, hierarchical learning). As a result only a part of the change is constructive, while its periphery (neighborhood) is destructive. This seems to be the case for change B in Fig. 6.3. As the bird was creating the Wiener entropy peak he also increased the Wiener entropy around the peak. This “collateral damage” shows up as a destructive change in Fig, 6.3 (the change labeled “B”).

There are other possible causes of destructive vocal changes. One such cause could be the attempt to segregate to neighboring parts of the syllable in order to control their development separately (the attempt to segment the action). The change labeled “C” in Fig. 6.3 could suggest such attempt. In such a case it may be preferable to first separate two parts of a syllable by “moving” them in opposite directions and only after the separation (when independent control of both parts becomes possible) change them in the right direction. This may be happening at label “D” in Fig. 6.3 where a narrow corrective change (red, constructive change) is happening.

The third possible cause of destructive changes may be neglect of some parts of the syllable when other parts are changing. This hypothesis suggests the existence of “motor attention”. It could be that some parts of the action that require a large portion of a limited resource of motor attention will cause other parts to slightly deteriorate. This could be particularly relevant when difficult parts of continuous actions are learned. The vocal change labeled “E” could be and example of such deterioration.

So far the data supporting these hypotheses are still anecdotal (observed in two out of eight birds studied). It would be interesting to perform an experiment in which young birds that can presumably only perform broad vocal changes would be reinforced to add a short time-scale element to their syllables as in Tumer and Brainard experiment with older birds (Tumer and Brainard, 2007). We could predict that early in development the birds will make broad changes with significant “collateral damage” in the periphery of the newly added element (such as label “A” in Fig. 6.3).

Chapter 7: Integration and Discussion

The central aim of this dissertation was to explore how complex, continuous actions can be learned effectively, given the conflict that exists between exploration and consolidation of different parts of the action. The solution to this problem is partitioning of the action into segments but while the advantages of such partitioning have been studied in theoretical models and machine learning algorithms (Doya, 2000), less is know about the implementation of this solution in animal learning models. Here we studied the learning of zebra finch syllables as a model of continuous action. We found evidence for partitioning of the syllables into segments to which exploratory variability could be applied independently. This implies that error is assessed locally, on the level of sub-syllabic structure. The second problem that arises in learning of continuous action by partition is how to determine the size of segments (granularity). Too many segments (high granularity) would present a burden to the memory, while too few segments (low granularity) would result in the original problem of the conflict between exploration and consolidation.

We will start this discussion with possible neuronal and peripheral mechanism that could be involved in local modification of the song. So the first question is at what time scales can the known neural substrates regulate exploratory variability.

6.1 Can AFP regulate exploratory variability in short intervals?

It has been shown that anterior frontal pathway (AFP), which is a likely source of exploratory variability (Olveczky, 2005, Andalman, and Fee, 2009), can modify the song locally (Kao et al., 2005).

Micro-stimulating the LMAN (the nucleus of AFP which projects directly to RA in the song production pathway) can trigger highly time-localized and acoustic feature-specific vocal changes in the song of adult birds. A stimulation of a particular *site* in LMAN can generate changes in specific parts of the song (always at the same parts). On the other hand, the modifications in one part of the song can be qualitatively different from those in another part. For example, micro-stimulations of the same site can increase the fundamental frequency in one syllable but decrease its magnitude in another syllable. Therefore, as this result suggests, the *site* of micro-stimulation in LMAN determines the time-positions of modifications but not necessarily their quality (the direction of modifications). On the other hand, stimulations in different sites in LMAN can have opposite effects in the same syllable. Therefore there are many sites in LMAN that can affect the same time-positions of the song but their effects on the song are qualitatively different.

This opens a question about how can LMAN induce exploration locally. If the same site always modifies the song features in a particular direction then many sites in LMAN should be activated at a particular time. This assembly of LMAN sites should only affect the same, short segment, of the song and thereby induce local exploration. In other words, different assemblies of neurons in LMAN should be assigned to different times in the song. Such organization is reminiscent of the

RA, where different assemblies of neurons are activated by the HVC at different times (Leonardo et al, 2005).

So far we have seen that AFP can modulate the song locally and that recruitment of different neuronal assemblies in LMAN could assign exploration to particular song segments. But whatever the mechanism of recruitment in LMAN is, for AFP to be able to adjust the gain of its output locally during singing, it needs to “know” the exact song time. This timing information could be provided to AFP via the HVC *Area X-projecting neurons.* Area X is the first nucleus in the AFP onto which HVC projects (see Figure 1.2 in Chapter 1). As mentioned in Chapter 1, the HVC produces a very sparse neuronal activity, which is extremely time-locked (Hahnloser et al, 2002). Although it has been shown that lesions to Area X, which should block this time information, have no obvious overall effect on song structure or on the magnitude of noise (Goldberg and Fee, 2011), it is possible nevertheless that short time-scale modulation in vocal exploration might be gated by Area X (Kojima and Doupe, 2009), or that AFP has other means of generating a song-time dependent signal.

6.2 Can differential sensitivity of RA neurons to AFP input regulate variability at short time scales?

Local regulation of exploratory variability could also be explained by differential sensitivity of RA pre-motor neurons to AFP input, across song elements. Each pre-motor RA neuron receives inputs from several HVC and AFP neurons. As mentioned above, different assemblies of RA neurons are activated at a particular time by HVC, which results in production of time dependent song elements. As a particular song element approaches the target the weight of HVC-RA synapses could increase, making the synapse less sensitive to the constant gain of variability coming from AFP. The AFP-RA connections are mediated by NMDA receptors, suggesting that the AFPs activity could modulate the HVC-RA synaptic weights (which are mediated by the mixture of AMPA and NMDA receptors).

Disambiguating between the two hypotheses (AFP or differential sensitivity of RA neurons) could be attempted by training birds with our AAAA->ABAB paradigm (Chapter 2) and testing the effect of micro-stimulation from AFP during singing of either syllable A or B, once the highly variable B syllable appears. Such a stimulation should cause brief vocal changes that are song-time specific as in the experiments by Kao et al. (Kao et al., 2005). The prediction is that if the sensitivity of RA neurons to AFP input is higher for the new syllable (B), such stimulation would have a stronger effect if delivered during performance of B. Alternatively, recording from LMAN should show increased activity or less inter-hemispheric synchrony (Wang et al., 2008) when singing B.

6.3 Reinforcement learning of continuous action without segmentation can be effective when the gain of exploratory variability is low

In Chapter 4 we discussed the age dependent reduction of plasticity and the decreasing gain of exploratory variability. As we saw a minimal amount of residual exploratory variability can be used in order to modify the song even at the old age, when song seems “crystallized” (Tumer and Brainard, 2007). This has been shown experimentally in the procedure where a negative reinforcement was used to induce a change the fundamental frequency of short song elements in older birds. We have noted in Chapter 4 that no evidence was found for the ability of birds to increase their exploratory variability when exposed to negative reinforcement in those experiments. Thus with the ageing the gain of variability seems to only decrease.

Minimal variability in older birds does not cause significant deterioration of the song. Consequently the local regulation of variability many no longer be necessary even when they need to modify very short time-scale song elements in the negative reinforcement experiments.

This finding is further supported by the study by Charlesworth et al in which negative reinforcement is used to effectively induce changes in particular song elements of older birds, *without specifying their time-position*, thereby providing only the global error (Charlesworth et al, 2011). Consider, for example, a song bout with five song elements (A, B, C, D and E). A negative reinforcer (punishment) may be introduced at the end of the song bout if the fundamental frequency of element B is too low. The subject bird will increase the average fundamental frequency (across many renditions) in element B during the training period. In this setting birds can learn to change even multiple song elements (in any direction specified by the experimentalists) with a single negative reinforcer (single global error estimate). Importantly, Charlesworth et al did not observe any increments of variability during learning that could otherwise be expected when a bird is trying to escape the negative reinforcer. Of course, with only global reinforcement provided the variability would have to increase along the entire trajectory of the continuous action. Therefore, one reason for why increased variability was not observed in those experiments could be that such global increment of exploration would cause more deterioration of song structure than in the case when error can be locally estimated. In any case, the result shows that segmentation of continuous action may not be necessary for learning.

Taken together the role of AFP is to provide exploratory variability signal to the motor pathway (Olveczky et al, 2005) and can maintain residual variability necessary to make short time-scale changes in older birds (Tumer and Brainard, 2007) even when the bird has no access to local error estimation (Charlesworth et al, 2011). Exploratory variability may also be used in a more "guided" way (Andalman and Fee, 2009) such that noise injected by AFP will bias the exploration in the direction of the target and there is strong evidence that AFP can affect short segments of the song (Kao et al, 2005). What our work adds to this body of knowledge is that in younger birds the exploratory variability can also be locally regulated either directly by AFP or the changes in sensitivity in RA to the AFP input. We have suggested experiments the electro-physiologists could perform in the future to further study the mechanisms of local regulation of vocal exploration. In Chapter 4 we also suggested that experiments with negative reinforcement be performed on younger birds where one could predict more exploration in the parts that need to change.

6.4 Hierarchical learning: increasing granularity of segmentation

As noted in the beginning of this dissertation (Chapter 1) there are two problems in learning of continuous action. The first problem arises from the conflict between exploration and consolidation. This problem can be solved by partitioning the action into segments that can then be learned independently from each other. But this leads us to the second problem, namely, how to determine the sizes of segments. As noted earlier, if the granularity is too high, the burden on the memory becomes too great and the local error assessment too demanding. If granularity is too low, then within the segments we run into the first problem again (conflict between exploration and consolidation).

The solution to this problem of granularity would be to gradually decrease the size of segments over the course of learning. This way high exploratory variability could be used across broad segments, early in the learning process, when consolidation has lower priority than exploration (because there is not much structure in place yet). Later in development, as some parts of the action are already close to the target (low error), consolidation of these parts becomes the greater priority and exploration becomes confined to only those parts of the action that still need to improve. We refer to this gradual decrement of the segment size (or increment of granularity) as *hierarchical learning*.

While we found some evidence for hierarchical learning in syllabic acoustic structure where syllables appear before the intra-syllabic events (see Chapter 4), a stronger support comes from the analysis of the respiratory pressure patterns (see Chapter 5).

We showed in Chapter 5 that during development of respiratory pressure pulses (EPs), fine temporal structure (high granularity) is added to the pre-existing coarse temporal structure (low granularity). Interestingly, this indication of hierarchical learning was only observed in tutored birds, while in birds that were never exposed to the normal song there was no indication of any increment of granularity. It is questionable if segmentation of the song ever takes place in untutored birds, even though, as discussed in Chapter 5, over several generations of cultural learning the number of song elements increases as they become shorter.

6.5 Does the partition of song on the level of articulation precede the partition on the acoustic level?

We have seen so far that there are neural substrates that could be involved in partitioning of the continuous action of singing. At what point of song development does the partitioning occur? There is some evidence that the first segmentation of the song into syllables may occur at the level of articulation before it happens in the acoustics (Veit et al., 2011). This study shows that the establishment of precisely timed gaps between expiratory pulses (EPs) in the sub-song may precede the creation of equally well timed gaps between the syllables. However, it is important to remember that syllables are already somewhat discrete units. If we are studying the segmentation of truly continuous actions, then sub-syllabic structure would be more relevant. Sub-song itself is also not a good representative of continuous action in need of segmentation because the gaps must be created anyways for respiratory demands. Therefore even if the timing in respiratory pressure becomes more precise before the timing of acoustic structure, this does not necessarily mean that the segmentation happens first on the level of articulation. In this respect we should look into segmentation within individual EPs (uninterrupted continuous action) and compare it to the acoustic structure. However, as discussed in Chapter 5, the acoustic output is the final product of integration of respiratory, syrengeal and beak activity. The imprecision of this integration may result in apparently delayed measurements of refinement on the acoustic level. This was the main motivation to compare activity in the syrinx itself (EMG) with the respiratory pressure. But as we saw in Chapter 5, there is no apparent lack of synchronicity between developmental changes in the pressure pattern and changes in syrengeal

EMG activity. Nor has a delay in development of fine acoustic structure behind articulation ever been observed.

In summary, it seems that the partitioning of continuous actions into segments that could be independently evaluated and explored has to happen on multiple levels of the song production. Partitioning of continuous action can improve learning as it resolves the conflict between exploration and consolidation. We have shown that exploratory variability can be regulated locally (Chapters 2 and 4). With learning other problems, such as co-articulatory constraints and “collateral damage” of vocal changes (suggested in Chapter 6) may arise. A possible solution to those problems can be hierarchical learning (presented in Chapter 5) where segmentation of continuous action becomes progressively finer over the course of development.

We hope that our findings will contribute to generating new hypotheses about neuronal mechanisms of learning of continuous actions that could be tested by electro-physiologists in birdsong as well as in other fields.

Bibliography

Andalman AS, Fee MS (2009) A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. Proc Nat Acad Sci USA 106:12518-12523.

Aronov D, Andalman AS, Fee MS (2008) A specialized forebrain circuit for vocal babbling in the juvenile songbird. Science 320:630-634.

Beers RJ (2009) Motor learning is optimally tuned to the properties of motor noise. *Neuron*, *63*(3), 406-17.

Bhushan N, Shadmehr R (1999) Computational nature of human adaptive control during learning of reaching movements in force fields. Biol Cybernetics 81: 39–60.
birdsong. J Neurobiol 36:441-453.

Boettiger CA, Doupe AJ (2001) Developmentally restricted synaptic plasticity in a songbird nucleus required for song learning. Neuron 31:809-818.

Bottjer SW, Miesner EA, Arnold AP (1984) Forebrain lesions disrupt development but not maintenance of song in passerine birds. Science 224:901-903.

Brainard MS, Doupe J (2000) Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. Nature 404:762-766.

Brainard MS, Doupe J (2001) Postlearning consolidation of birdsong: stabilizing effects of age and anterior forebrain lesions. J Neurosci 21:2501-2517.

Calder WA (1970) Respiration during song in the canary (Serinus canaria). Comp Biochem Physiol 32:251-258

Charlesworth JD, Tumer EC, Warren TL, Brainard MS (2011) Learning the microstructure of successful behavior Nature neuroscience, 14(3), 373-380.

Conrad B, Schönle P (1979) Speech and respiration. European Archives of Psychiatry and Clinical Neuroscience Volume 226, Number 4, 251-268

Cooper BG, Goller F (2004) Partial muting leads to age-dependent modification of motor patterns underlying crystallized zebra finch song. J Neurobiol 61:317-332.

Cynx, J (1990) Experimental determination of a unit of song production in the zebra finch (Taeniopygia guttata). Journal of Comparative Psychology, Vol 104(1)

Derégnaucourt S, Mitra PP, Fehér O, Pytte C, Tchernichovski O (2005) How sleep affects the developmental learning of bird song. Nature 433:710-716.

Doupe AJ, Solis MM, Kimpo R (2004) Cellular, circuit, and synaptic mechanisms in song learning. Ann NY Acad Sci 1016: 495-523.

Doya K (2000) Reinforcement learning in continuous time and space. Neural Comput 12:219-245.

Du P and Troyer TW (2006) A Segmentation Algorithm for Zebra Finch Song at the Note Level. Neurocomputing, 69(10-12): 1375-1379

Engel KC, Soechting JF (2011) Manual Tracking in Two Dimensions Manual Tracking in Two Dimensions. *Journal of Neurophysiology*: 3483-3496.

Ferster CB, Skinner BF, (1957) Schedules of Reinforcement. East Norwalk, CT, US: Appleton-Century-Crofts vii, 744 pp.

Fiete IR, Fee MS, Seung HS (2007) Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. J neurophysiol 98:2038-2057.

Goldberg JH, Fee MS (2011) Vocal babbling in songbirds requires the basal ganglia-recipient motor thalamus but not the basal ganglia. J Neurophysiol 105:2729-2739.

Goller F (2002) Respiratory units of motor production and song imitation in the zebra finch. J Neurobiol 51:129-141.

Goller F, Cooper BG (2004) Peripheral motor dynamics of song production in the zebra finch. Annals NY Acad Sci 1016:130-152.

Goller F, Daley MA (2001) Novel motor gestures for phonation during inspiration enhance the acoustic complexity of birdsong. Proc R Soc Lond B 268:2301–2305.

Goller F, Mallinckrodt MJ, Torti SD (2004) Beak gape dynamics during song in the zebra finch. J Neurobiol 59:289-303.

Goller F, Suthers RA (1996) Role of syringeal muscles in controlling the phonology of bird song. J Neurophysiol 76:287-300.

Goller F, Suthers RA (1999) Bilaterally Symmetrical Respiratory Activity during Lateralized Birdsong. J Neurobiol 41(4):513-23.

Haesler S, Rochefort C, Georgi B, Licznerski P, Osten P, Scharff, C (2007) Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. PLoS Biol 5(12):e321.

Immelmann K (1969) Song development in the zebra finch and other estrilid finches. In: Bird vocalizations (Hinde RA, ed), pp 61-74. New York: Cambridge UP.

Jarvis ED, Scharff C, Grossman, MR, Ramos J, Nottebohm F (1998) For whom the bird sings: context-dependent gene expression Neuron, 21(4), 775-88.

Kaelbling LP, Littman ML, Moore AW (1996) Reinforcement learning: A Survey Journal of Artificial Intelligence Research 4: 237–285.

Kao MH, Doupe AJ, Brainard MS (2005) Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. Nature 433:638-643.

Kawato M (1999) Internal models for motor control and trajectory planning Curr Op Neurobiol 9: 718–727.

Knudsen EI, Knudsen PF (1989) Vision calibrates sound localization in developing barn owls. J Neurosci 9:3306-3313

Konishi M (1965) The role of auditory feedback in the control of vocalization in the white-crowned sparrow. Z Tierpsychol 22(7):770-83.

Kozhevnikov AA, Fee MS (2007) Singing-related activity of identified HVC neurons in the zebra finch. J Neurophysiol 97:4271-4283.

Kubke MF, Yazaki-Sugiyama Y, Mooney R, Wild JM (2005) Physiology of neuronal subtypes in the respiratory–vocal integration nucleus retroamigualis of the male zebra finch. J Neurophysiol 94:2379–2390.

Lipkind D, Tchernichovski O (2011) Quantification of Behavior Sackler Colloquium: Quantification of developmental birdsong learning from the subsyllabic scale to cultural evolution. Proc Nat Acad Sci USA 2011:1-8.

Marler P (1970) A comparative approach to vocal learning song development in white-crowned sparrows. Journal of Comparative Physiological Psychology Monographs 71, pp. 1–25.

Marler P, Sherman V (1985) Innate differences in singing behavior of sparrows reared in isolation from adult conspecific song. Anim Behav 33:57–71.

Méndez JM, Dall'Asén AG, Cooper BG, Goller F (2010) Acquisition of an acoustic template leads to refinement of song motor gestures. J Neurophysiol 104:984-993.

Miller JE, Hilliard AT, White S (2010) Song practice promotes acute vocal variability at a key stage of sensorimotor learning. PloS One 5(1):e8592.

Morasso P (1981) Spatial control of arm movements. Exp Brain Res 42: 223–227.

Morrison RG, Nottebohm FJ (1993) Role of a telencephalic nucleus in the delayed song learning of socially isolated zebra finches. J Neurobiol 24:1045-1064.

Müller H, Sternad D (2009) Motor learning: changes in the structure of variability in a redundant task. In: Progress in motor control: a multidisciplinary perspective (Sternad D, ed), pp 439-456. New York: Springer.

Olveczky BP, Andalman AS, Fee MS (2005) Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. PLoS Biol 3(5):e153.

Saar S, Mitra PP, Derégnaucourt S, Tchernichovski O (2007) Developmental song learning in the zebra finch. Invited paper by the Annals of the New York Academy of Science to the 2007 issue of Neurobiology of Birdsong

Scharff C, Nottebohm F (1991) A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. J Neurosci 11:2896-2913.

Shadmehr R, Mussa-Ivaldi FA (1994) Adaptive representation of dynamics during learning of a motor task. Journal of Neuroscience 14: 3208–3224.

Skinner BF (1938) The Behavior of Organisms: An Experimental Analysis. ISBN 1-58390-007-1

Slater PJB (1983) Bird song learning: theme and variations. Perspectives in ornithology (Brush AH and GA Clark, eds), pp 475-499. Cambridge, England: Cambridge UP.

Smart WD, Kaelbling LP (1910) Practical Reinforcement Learning in Continuous Spaces Artificial Intelligence.

Sundberg Johan (1993) Breathing behavior during singing. The NATS Journal 49: 2–9, 49–51. ISSN 0884-8106

Suthers RA (1997) Peripheral control and lateralization of bird-song. J Neurobiol 33:632-652.

Suthers RA, Goller F, Pytte C (1999) The neuromuscular control of birdsong. Philos Trans R Soc Lond B Biol Sci 354(1385):927–939.

Suthers RA, Margoliash D (2002) Motor control of birdsong. Current Opinion in Neurobiology 12:684-690.

Suthers RA, Vallet E, Tanvez A, Kreutzer M (2004) Bilateral song production in domestic canaries. J Neurobiol 60:381-393.

Suthers RA, Zollinger SA (2004) Producing song: The vocal apparatus. In: Behavioral Neurobiology of Bird Song. (Zeigler HP, Marler P, eds) Ann NY Acad Sci 1016:109-29.

Tchernichovski O, Lints T, Mitra PP, Nottebohm F (1999) Vocal imitation in zebra finches is inversely related to model abundance. Proc Nat Acad Sci USA 96:12901-12904.

Tchernichovski O, Mitra PP (2004) Sound Analysis Pro user manual. Available at: http://ofer.sci.ccny.cuny.edu.

Tchernichovski O, Mitra PP, Lints T, Nottebohm F (2001) Dynamics of the vocal imitation process: how a zebra finch learns its song. Science 291(5513):2564-2569.

Tchernichovski O, Nottebohm F, Ho C, Pesaran B, Mitra P (2000) A procedure for an automated measurement of song similarity. Anim Behav 59:1167-1176.

Tumer EC, Brainard MS (2007) Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. Nature 450:1240-1244.

Veit L, Aronov D, Fee MS (2011) Learning to breathe and sing: Development of respiratory-vocal coordination in young songbirds Journal of neurophysiology. *In press*

Vicario DS Organization of the zebra finch song control system: II. Functional organization of outputs from nucleus Robustus archistriatalis (1991) J of Comperative Neurobiology 22;309(4):486-94.

Wang CZ, Herbst Ja, Keller GB, Hahnloser RH (2008) Rapid interhemispheric switching during vocal production in a songbird. PLoS biology 6(10):e250.

Watson JB (1913) Psychology as the Behaviorist Views it. First published in Psychological Review, 20, 158-17

Wild JM, Goller F and Suthers (1998) Inspiratory muscle activity during bird song. Journal of neurobiology, 36(3), 441-53.

Wild JM, Williams MN, Suthers RA (2000) Neural pathways for bilateral vocal control in songbirds**.** J Comp Neurol 423:413-426.