G. B. Mindlin
R. Laje

The Physics of Birdsong

Springer

BIOLOGICAL AND MEDICAL PHYSICS
BIOMEDICAL ENGINEERING

BIOLOGICAL AND MEDICAL PHYSICS, BIOMEDICAL ENGINEERING

BIOLOGICAL AND MEDICAL PHYSICS, BIOMEDICAL ENGINEERING

The fields of biological and medical physics and biomedical engineering are broad, multidisciplinary and dynamic. They lie at the crossroads of frontier research in physics, biology, chemistry, and medicine. The Biological and Medical Physics, Biomedical Engineering Series is intended to be comprehensive, covering a broad range of topics important to the study of the physical, chemical and biological sciences. Its goal is to provide scientists and engineers with textbooks, monographs, and reference works to address the growing need for information.

Books in the series emphasize established and emergent areas of science including molecular, membrane, and mathematical biophysics; photosynthetic energy harvesting and conversion; information processing; physical principles of genetics; sensory communications; automata networks, neural networks, and cellular automata. Equally important will be coverage of applied aspects of biological and medical physics and biomedical engineering such as molecular electronic components and devices, biosensors, medicine, imaging, physical principles of renewable energy production, advanced prostheses, and environmental control and engineering.

Gabriel B. Mindlin Rodrigo Laje

The Physics of Birdsong

With 66 Figures

Springer

Prof.Dr. Gabriel B. Mindlin
Rodrigo Laje
Universidad de Buenos Aires
FCEyN
Departamento de Física
Pabellón I, Ciudad Universitaria
C1428EGA Buenos Aires
Argentina
e-mail: gabo@df.uba.ar
rodrigo@df.uba.ar

Library of Congress Control Number: 2005926242

ISSN 1618-7210

ISBN-10 3-540-25399-8

ISBN-13 978-3-540-25399-0

Springer is a part of Springer Science+Business Media
springeronline.com

Printed in The Netherlands

Cover concept by eStudio Calamar Steinen using a background picture from The Protein Databank (1 Kzu). Courtesy of Dr. Antoine M. van Oijen, Department of Molecular Physics, Huygens Laboratory, Leiden University, The Netherlands. Reprinted with permission from Science 285 (1999) 400–402 ("Unraveling the Electronic Structure of Individual Photosynthetic Pigment-Protein Complexes", by A.M. van Oijen et al.) Copyright 1999, American Association for the Advancement of Science.

Typesetting by the authors and Techbooks using a Springer LaTeX package
Cover production: *design & production* GmbH, Heidelberg

Printed on acid-free paper SPIN 11326915 57/3141/jvg - 5 4 3 2 1 0

Preface

Few sounds in nature show the beauty, diversity and structure that we find in birdsong. The song produced by a bird that is frequently found in the place where we grew up has an immense evocative power, hardly comparable with any other natural phenomenon. These reasons would have been more than enough to attract our interest to the point of working on an aspect of this phenomenon. However, in recent years birdsong has also turned into an extremely interesting problem for the scientific community. The reason is that, out of the approximately 10 000 species of birds known to exist, some 4000 share with humans (and just a few other examples in the animal kingdom) a remarkable feature: the acquisition of vocalization requires a certain degree of exposure to a tutor. These vocal learners are the oscine songbirds, together with the parrots and hummingbirds. For this reason, hundreds of studies have focused on localizing, within the birds' brains, the regions involved in the learning and production of the song. The hope is to understand through this example the mechanisms involved in the acquisition of a general complex behavior through learning. The shared, unspoken dream is to learn something about the way in which we humans learn speech. Studies of the roles of hormones, genetics and experience in the configuration of the neural architecture needed to execute the complex task of singing have kept hundreds of scientists busy in recent years.

Between the complex neural architecture generating the basic instructions and the beautiful phenomenon that we enjoy frequently at dawn stands a delicate apparatus that the bird must control with incredible precision. This book deals with the physical mechanisms at play in the production of birdsong. It is organized around an analysis of the song "up" toward the brain. We begin with a brief introduction to the physics of sound, discussing how to describe it and how to generate it. With these elements, we discuss the avian vocal organ of birds, and how to control it in order to produce different sounds. Different species have anatomically different vocal organs; we concentrate on the case of the songbirds for the reason mentioned above. We briefly discuss some aspects of the neural architecture needed to control the vocal organ, but our focus is on the physics involved in the generation of the song. We discuss some complex acoustic features present in the song that are generated when simple neural instructions drive the highly complex

vocal organ. This is a beautiful example of how the study of the brain and physics complement each other: the study of neural instructions alone does not prepare us for the complexity that arises when these instructions interact with the avian vocal organ.

This book summarizes part of our work in this field. At various points, we have interacted with colleagues and friends whom we would like to thank. In the first place, Tim J. Gardner, who has shared with us the first, exciting steps of this research. At various stages of our work in the field, we had the privilege of working with Guillermo Cecchi, Marcelo Magnasco, Marcos Trevisan, Manuel Eguía and Franz Goller, who are colleagues and friends. The influence of several discussions with other colleagues has not been minor: Silvina Ponce Dawson, Pablo Tubaro, Juan Pablo Paz, Ale Yacomotti, Ramón Huerta, Oscar Martínez, Guillermo Dussel, Lidia Szczupak, Henry Abarbanel, Jorge Tredicce, Pablo Jercog and Héctor Mancini. The support of Fundación Antorchas, Universidad de Buenos Aires, CONICET and ANPCyT has been continuous. Several recordings were performed in the E.C.A.S. Villa Elisa nature reserve in Argentina, with the continuous support of its staff. Part of this book was written during a period in which Gabriel Mindlin enjoyed the hospitality of the Institute for Nonlinear Science, University of California at San Diego. Heide Doss-Hammel patiently edited the first version of this manuscript, and enriched it with her comments.

One of us (R. L.) thanks Laura Estrada, and Jimena, Santiago, Pablo and Kanky, for their continuous support and love.

Finally, it was the support of Silvia Loza Montaña, Julia and Iván that kept this project alive through the difficult moments in which it was conceived.

Buenos Aires,
April 2005

Gabriel B. Mindlin
Rodrigo Laje

Contents

1 Elements of the Description

There is a wide range of physical phenomena behind birdsong. Physics allows us to understand what mechanisms are used in order to generate the song, what parameters must be controlled, and what part of the complexity of the sound is the result of the physics involved in its generation. The understanding of these processes will take us on a journey in which we shall visit classical mechanics, the theory of fluids [Landau and Lifshitz 1991, Feynman et al. 1970], and even some modern areas of physics such as nonlinear dynamics [Solari et al. 1996]. Ultimately, all these processes will be related to the sounds of birdsong described in this text. For this reason, it is appropriate to begin with a qualitative description of sound. Even if it is likely that the reader is familiar with the concepts being discussed, this will allow us to establish definitions of some elements that will be useful in our description and analysis of birdsong.

1.1 Sound

1.1.1 A Metaphor

Let us imagine a group of people standing in line, with a small distance between each other. Let us assume that the last person in the line tumbles and, in order to avoid falling, extends his/her arms, pushing forward the person in front. This person, in turn, reacts just like the person that pushed him/her: in order to avoid falling, this person pushes the person in front, and so on. None of the people in the line undergoes a net displacement, since every person has reacted by pushing someone else, and returning immediately to their original position. However, the "push" does propagate from the end of the line to the beginning. In fact, the first person in line can also try to avoid falling, by pushing some object in front of him/her. In other words, he/she can do work if the object moves after the push. It is important to realize that the propagation of this "push" along the line occurs thanks to *local* displacements of each of the persons in the line: each person moves just a small distance around their original position although the "push" propagates all along the line.

Maybe the person that began this process finds the spectacle of a propagating "push" amusing, and repeats it from time to time (trying to implement this thought experiment is not the best idea). The time between "pushes" is what we call the *period* of the *perturbation*. A related concept is the *frequency*: the number of pushes per unit of time (for example, "pushes" per second). Within our metaphor, each subject can either experience a slight deviation from his/her position of equilibrium or be close to falling. The quantity that describes the size of the perturbation is called its *amplitude*.

From this metaphor, we can extract the fact then that it is possible to propagate energy (capacity to do work) through a medium (a group of people standing in line) that undergoes perturbations on a local scale (no one moves far away from their equilibrium position), owing to a generator of perturbations (the last person in line, the one with a curious sense of humor) that produces a signal (a sequence of pushes) of a given amplitude and at a certain frequency.

1.1.2 Getting Serious

While it is true that metaphors can help us construct a bridge between a phenomenon close to our experience and another one which requires indirect inferences, it is also true that holding on to them for too long can hinder us in our understanding of nature. Sound is a phenomenon of propagative character, as in the situation described before. But an adequate description of the physics involved must consider carefully the properties of the real propagative medium, which, in the present case of interest, is air.

If an object moves slowly in air, a smooth flow is established around it. If the movement is so fast that such a flow cannot be established, compression of the air in the vicinity of the moving object takes place, causing a local change in pressure. In this way, we can originate a propagative phenomenon like the one described in our metaphor. In order to establish sound, the excess pressure must be able to push the air molecules in its vicinity (in terms of our metaphor, the people in the line should not be more than approximately an arm's length away from each other). Can we state a similar condition for the propagation of sound in air?

As opposed to what happens in our metaphor, the molecules of air are not static, or in line. On the contrary, they are moving and colliding with each other in a most disorderly manner, traveling freely during the time intervals between successive collisions. The average distance of travel between collisions is known as the *mean free path*. Therefore, if we establish a high density of molecules in a region of space, the escaping particles will push the molecules in the region of low density only if the density varies noticeably over distances greater than the mean free path. If this is not the case, the region of high density will "smoothen" without affecting its vicinity. For this reason, the description of sound is given in terms of the behavior of "small portions of air" and not of individual molecules. Here is an important

difference between our metaphor and the description of sound. The proper variables to describe the problem will be the density (or pressure) and velocity of the small portion of air, and not the positions and velocities of individual molecules [Feynman et al. 1970].

1.1.3 Sound as a Physical Phenomenon

The physics of sound involves the motion of some quantity of gas in such a way that local changes of density occur, and that these changes of density lead to changes in pressure. These nonuniform pressures are responsible for generating, in turn, local motions of portions of the gas.

In order to describe what happens when a density perturbation is generated, let us concentrate on a small portion of air (small, but large enough to contain many molecules). We can imagine a small cube of size Δx, and our portion of air enclosed in this imaginary volume. Before the sound phenomenon is established, the air is at a given pressure P_0, and the density ρ_0 is constant (in fact, the value of the pressure is a function of the value of the density). Before a perturbation of the density is introduced, the forces acting on each face of the cube are equal, since the pressure is uniform. Therefore, our portion of air will be in equilibrium. We insist on the following: when we speak about a small cube, we are dealing with distances larger than the mean free path. Therefore, the equilibrium that we are referring to is of a macroscopic nature; on a small scale with respect to the size of our imaginary cube, the particles move, collide, etc.

Now it is time to introduce a kinematic perturbation of the air in our small cube, which will be responsible for the creation of a density perturbation ρ_e. We do this in the following way: we displace the air close to one of the faces at a position x by a certain amount $D(x,t)$ (in the direction perpendicular to the face), and the rest of the air is also displaced in the same direction, but by a decreasing amount, as in Fig. 1.1. That is, the air at a position $x+\Delta x$ is displaced by an amount $D(x+\Delta x,t)$, which is slightly less than $D(x,t)$. As the result of this procedure, the air in our imaginary cube will be found in a volume that is compressed, and displaced in some direction. We now have a density perturbation ρ_e. Conservation of mass in our imaginary cube (that is, mass before displacement = mass after displacement) leads us to

$$\begin{aligned}\rho_0\,\Delta x &= (\rho_0+\rho_e)\,[(x+\Delta x+D(x+\Delta x,t))-(x+D(x,t))]\\ &= (\rho_0+\rho_e)\left(\Delta x+\frac{\partial D}{\partial x}\Delta x\right)\\ &= \rho_0\,\Delta x+\rho_0\frac{\partial D}{\partial x}\Delta x+\rho_e\,\Delta x+\rho_e\frac{\partial D}{\partial x}\Delta x\,. \end{aligned} \tag{1.1}$$

Let us keep only the linear terms by throwing away the term containing $\rho_e\,\partial D/\partial x$ as a second-order correction, since we can make the displacement

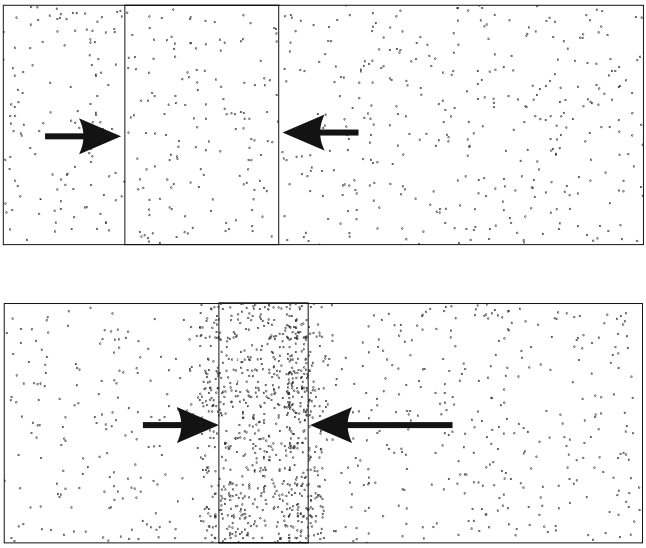

Fig. 1.1. Propagation of air density perturbations. (*top*) The air in a small imaginary cube is initially in equilibrium. We now "push" from the left, displacing the left face of the imaginary cube and compressing the air in the cube. (*bottom*) A density perturbation is created by the push, leading to an imbalance of forces in the cube. The forces now try to restore the air in the cube to its original position. At the same time, however, the portion of air in the "next" cube will be pushed in the same direction as the first portion was, propagating the perturbation

and hence the density perturbation as small as we want. Solving for ρ_e, (1.1) now reads

$$\rho_e = -\rho_0 \frac{\partial D}{\partial x} \,. \tag{1.2}$$

By virtue of the way we have chosen to displace the air (a decreasing displacement), air has accumulated within the cube, which means that we have created a positive density fluctuation.

What can we say about the dynamics of the problem now? Since we have created a nonuniform (and increasing) density in the direction of the displacements, we have established an increasing pressure in the same direction. By doing this, we have broken the equilibrium of forces acting on our portion of air. We have moved the faces, but by doing so, we have created an imbalance of density and pressure that tries to take our portion of air back to its original position, in a restitutive way. Another consequence is seen in the fate of a second portion of air, close to the original one in the direction in which we generated the compression. The imbalance of pressures around the new portion of air will lead to new displacements in the direction in which we generated our original perturbation, as shown in Fig. 1.1: a picture that does not differ much from the propagation of "pushes" discussed before.

With the help of Newton's laws for the air in our original imaginary cube, and ignoring the effects of viscosity, the restitutive effect of this imbalance may be written as follows:

$$\begin{aligned}\rho_0 \,\Delta x \,\frac{\partial^2 D}{\partial t^2} &= -\left[P(x+\Delta x,t) - P(x,t)\right] \\ &= -\frac{\partial p}{\partial x}\Delta x\,,\end{aligned} \tag{1.3}$$

where $P = P_0 + p$ is the pressure and p is the (nonuniform) pressure perturbation, or *acoustic pressure*. In addition, assuming that the pressure perturbations are linear functions of the density perturbations (which holds if the density perturbations are small enough), we can write the equation of state

$$p = \frac{\kappa}{\rho_0}\,\rho_e\,, \tag{1.4}$$

where κ is the adiabatic bulk modulus.

So far, we have a conservation law (1.2), a force law (1.3) and an equation of state (1.4). With these ingredients, we can write an equation for p only. If we differentiate (1.2) twice with respect to t, we obtain

$$\frac{\partial^2 \rho_e}{\partial t^2} = -\rho_0 \frac{\partial^2}{\partial t^2}\frac{\partial D}{\partial x}\,. \tag{1.5}$$

On the other hand, the differentiation of (1.3) with respect to x gives us

$$\rho_0 \frac{\partial}{\partial x}\frac{\partial^2 D}{\partial t^2} = -\frac{\partial^2 p}{\partial x^2}\,. \tag{1.6}$$

Writing ρ_e in terms of p and equating both expressions, we obtain the acoustic wave equation

$$\frac{\partial^2 p}{\partial t^2} = c^2\,\frac{\partial^2 p}{\partial x^2}\,, \tag{1.7}$$

where $c = \sqrt{\kappa/\rho_0}$ is the speed of sound, which is 343 m/s in air at a temperature of 20°C and atmospheric pressure. This is the simplest equation describing sound propagation in fluids. Some assumptions have been made (namely, sound propagation is lossless and the acoustic disturbances are small), and the reader may feel suspicious about them. However, excellent agreement with experiments on most acoustic processes supports this lossless, linearized theory of sound propagation. It is interesting to notice that the same equation governs the behavior of the variable D (displacement) and the particle velocity $v = -\partial D/\partial t$.

1.1.4 Sound Waves

Sound waves are constantly hitting our eardrums. They arrive in the form of a constant perturbation (such as the buzz of an old light tube) or a sudden

shock (such as a clap); they can have a pitch (such as a canary song) or not (such as the wind whispering through the trees). Sound waves can even seem to be localized in space, as in the "hot spots" that occur when we sing in our bathroom: sound appears and disappears according to our location.

What is a sound wave? It is the propagation of a pressure perturbation (in much the same way as a push propagates along a line). Mathematically speaking, a sound wave is a solution to the acoustic wave equation. By this we mean a function $p = p(x, t)$ satisfying (1.7). Every sound wave referred to in the paragraph above can be described mathematically by an appropriate solution to the acoustic wave equation. The buzz of a light tube or a note sung by a canary, for instance, can be described by a *traveling wave*. What is the mathematical representation of such a wave? Let us analyze a spatiotemporal function of space and time of the following form:

$$p(x, t) = p(x - ct)\,. \tag{1.8}$$

If we call the difference $x - ct = u$, then it is easy to see that taking the time derivative of the function twice is equivalent to taking the space derivative twice and multiplying by c^2. The reason is that $\partial p/\partial x = dp/du$, while $\partial p/\partial t = -cdp/du$. In other words, a function of the form (1.8) will satisfy the equation (1.7). Interestingly enough, it represents a traveling disturbance. We can visualize this in the following way: let us take a "picture" of the spatial disturbances of the problem by computing $p_0 = p(x, 0)$. The picture will look exactly like a picture taken at $t = t^*$, if we displace it a distance $x^* = ct^*$. It is interesting to notice that just as a function of the form (1.8) satisfies the wave equation, a function of the form $p(x, t) = p(x + ct)$ will also satisfy it. In other words, waves traveling in both directions are possible results of the physical processes described above. Maybe even more interestingly, since the wave equation (1.8) is linear, a sum of solutions is a possible solution. The spatiotemporal patterns resulting from adding such counterpropagating traveling waves are very interesting, and can be used to describe phenomena such as the "hot spots" in the bathroom. They are called "standing waves" and will be discussed as we review some elements that are useful for their description.

1.1.5 Detecting Sound

To detect sound, we need somehow to measure the pressure fluctuations. One way to do this is to use a microphone, which is capable of converting pressure fluctuations into voltages. Now we are able to analyze Fig. 1.2, which is a typical display of a record of a sound. The sound wave, that is, the propagation of a pressure perturbation, reaches our microphone and moves a mechanical part. This movement induces voltages in a circuit, which are recorded. In Fig. 1.2, we have plotted the voltage measured (which is proportional to the pressure of the sound wave in the vicinity of the microphone) as a function of

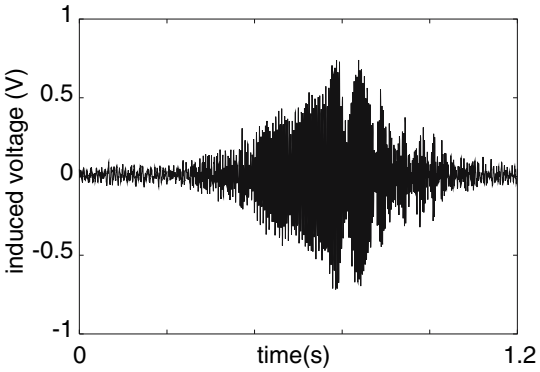

Fig. 1.2. A sound wave, as recorded by a microphone. A mechanical part within the microphone (for instance, a membrane or a piezoelectric crystal, capable of sensing tiny air vibrations) moves when the sound pressure perturbation reaches the microphone. The movement of this mechanical part induces a voltage in the microphone's circuit, which is recorded as a function of time. In this zoomed-out view of the recording, we can see hardly any details of the oscillation; instead, however, we could certainly draw the "envelope", which is a measure of how the sound amplitude changes with time

time. In this way, we can visualize how the pressure in the vicinity of the microphone varies as the recording takes place. In this figure, we have displayed 52 972 voltage values separated by time intervals of 1/44 100 s (i.e., a total recording time of 1.2 s). The inverse of this discrete interval of time is known as the *sampling frequency*, in this case 44 100 Hz. The larger the sampling frequency, the larger the number of data points representing the same total time of recording, and therefore the better the quality. This record corresponds to the song of the great grebe (*Podiceps major*) [Straneck 1990a].

1.2 Frequency and Amplitude

1.2.1 Periodic Signals vs. Noise

We now have the elements that we need to move forward and to present other elements important for the description of sound records. A sound source produces a signal that propagates in the air, generating pressure perturbations in the vicinity of a microphone. What do the time records of different sounds look like? In Fig. 1.3, we have two records corresponding to different sounds. The first one corresponds to what we call "noise" (for example, we might record the sound of the wind while we wait for the song of our favorite bird). The second record corresponds to what we would identify as a "note", a sound with a given and well-determined frequency. In fact, this record corresponds to a fraction of the great grebe's song (3/1000) s long. The first characteristic

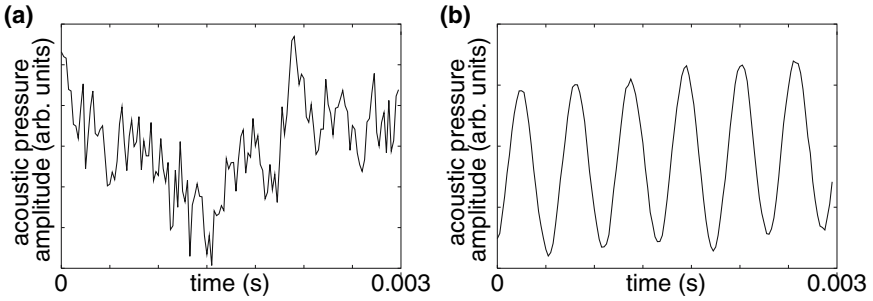

Fig. 1.3. Noise vs. pitched sound waves. (**a**) A very irregular sound wave (here, the wind recorded in the field) is what we call "noise". (**b**) In contrast, when the sound wave is regular or periodic (such as the fraction of the great grebe's song shown here), our ear is able to recognize a pitch, and we call it a "note"

that emerges from a comparison between the two records is the existence of a regularity in the second one. This record is almost *periodic*, i.e., it has similar values at regular intervals of time. This periodicity is recognized by our ear as a pure note. In contrast, when the sound is extremely irregular, we call it noise.

Let us describe pure notes. The periodicity of a signal in time allows us to give a quantitative description of it: we can measure its *period* T (the time it takes for a signal to repeat itself) or its *frequency* f, that is, the inverse of the period. The frequency represents the number of oscillations per unit of time, and is related to the parameter ω (called the *angular frequency*) through $\omega = 2\pi f$. If time is measured in seconds, the unit of frequency is known as the hertz ($1\,\mathrm{Hz} = 1/\mathrm{s}$). What does this mean in terms of something more familiar? Simply how high or low the pitch is. The higher the frequency, the higher the pitch.

Let us assume that the pure note corresponds to a traveling wave. In this case, the periodicity in time leads to a periodicity in space. For this reason, one can define a *wavelength* in much the same way as we defined a period for the periodicity in time. The meaning of the wavelength λ is easily seen by taking an imaginary snapshot of the sound signal and measuring the distance between two consecutive crests. It has, of course, units of distance such as meters or centimeters. A related parameter is the *wavenumber* $k = 2\pi/\lambda$. The wavenumber and angular frequency (and therefore the wavelength and frequency) are not independent parameters; they are related through

$$\omega = ck\,, \tag{1.9}$$

where c is the only parameter appearing in the wave equation (1.7), that is, the sound velocity.

1.2.2 Intensity of Sound

In the previous section, we were able to define the units of the period and the frequency. Now that we have a description of the nature of the sound perturbation, we shall concentrate on its *amplitude*. For a periodic wave such as the one displayed in Fig. 1.3b, the amplitude is the number that measures the maximum value of the departure from the average of the oscillating quantity.

Since, for a gas, the pressure is a function of the density, we can perform a description of the sound in terms of the fluctuations of either quantity. Traditionally, the option chosen is to use the pressure. Therefore, we have to describe how much the pressure P varies with respect to the atmospheric pressure P_0 when a sound wave arrives. Let us call this pressure p (that is, the increment of pressure when the sound wave arrives, with respect to the atmospheric value), and its amplitude A. Now, the minimum value of this quantity that we can hear is tiny: only 0.00000000019 times atmospheric pressure. Let us call this the reference pressure amplitude A_{ref}. We can therefore measure the intensity of a sound as the ratio between the sound pressure amplitude when the wave arrives, A, and the reference pressure amplitude A_{ref}.

This strategy is the one used to define the units of sound intensity. However, since the human ear has a *logarithmic* sensitivity (that is, it is much more sensitive at lower intensities), the sound intensity is measured in *decibels* (dB), which indicate how strong a pressure fluctuation with respect to a reference pressure is, but the intensity is measured in a way that reflects this way of perceiving sound. The sound pressure level I is therefore defined as

$$I = 20 \log_{10}(A/A_{\text{ref}})\,. \tag{1.10}$$

A sound of 20 dB is 10 times as more intense (in pressure values) as the weakest sound that we can perceive, while a sound of 120 dB (at the threshold of pain) is a million times as intense.

In Fig. 1.4, we show a series of familiar situations, indicating their characteristic frequencies and intensities. For example, a normal conversation has a typical intensity of 65 dB, and a rock concert can reach 115 dB (close to the sound intensity of an airplane taking off at a distance of a few meters, and close to the pain threshold). In terms of frequencies, the figure begins close to 20 Hz, the audibility threshold for humans. Close to 500 Hz, we place a note sung by a baritone, while at 6000 Hz we locate a tonal sound produced by a canary.

1.3 Harmonics and Superposition

1.3.1 Beyond Frequency and Amplitude: Timbre

We can tell an instrument apart from a voice, even if both are producing the same note. What is the difference between these two sounds? We need

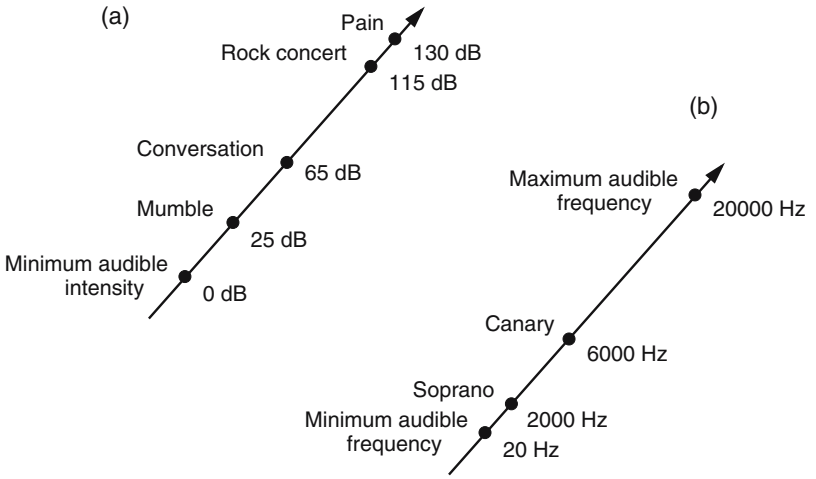

Fig. 1.4. Intensity and frequency ranges for the human ear. (**a**) The intensity scale starts at 0 dB, which does not mean the absence of sound but is the minimum intensity for a sound to be audible. A sound of 130 dB or more (the pain threshold) can cause permanent damage to the ear even if the exposure is short. (**b**) The minimum frequency of a pure sound for which our ear can recognize a pitch is around 20 Hz, that is, a wave oscillating only 20 times per second. The highest audible frequency for humans is around 20 000 Hz, although this depends on age, for instance. Unlike bats and dogs, birds cannot hear frequencies beyond the human limit (known as ultrasonic frequencies)

more than the period and the intensity to describe a sound. What is missing? What do we need in order to describe the *timbre*?

According to our description, the pitch of a note depends on the time it takes for the sound signal to repeat itself, i.e. the period T. But a signal can repeat itself without being as simple as the one displayed in Fig. 1.3b. In Fig. 1.5 (top curve), we show a sound signal corresponding to the same note as in Fig. 1.3b. The period T is indeed the same, but the signal displayed in Fig. 1.5 looks more complex. It is not a simple oscillation, and in fact we show in the figure that the signal is the sum of two simple oscillations. The first of these has the same period as the note itself. The second signal has a smaller period (in this case, precisely half the period of the note). If a signal repeats itself after a time $T/2$, it will also repeat itself after a time T. Therefore, the smallest time after which the complex signal will repeat itself is T. Our composite note will have a period T, as in the signal displayed in Fig. 1.3b, but it will sound different. The argument does not restrict us to adding two simple signals. We could keep on adding components of period T/n, where n is any integer, and still have a note of period T. The lowest frequency in this composite signal is called the *fundamental frequency* $F_1 = 1/T$, and the components of smaller period with frequencies $F_2 = 2/T$, $F_3 = 3/T, \ldots$,

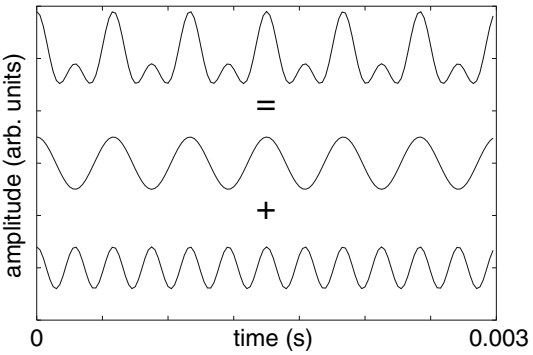

Fig. 1.5. Components of a complex oscillation. The sound wave at the top is not a simple or *pure* oscillation. Instead, it is the sum of two simple oscillations called its *components*, shown below. The components of a complex sound are usually enumerated in order of decreasing period (or increasing frequency): the first component is the one with the largest period of all the components, the second component is the one with the second largest period, and so on. Note that the period of the complex sound is equal to the period of its first component. The frequency of the first component is also called the *fundamental frequency*

$F_n = n/T$ are called the *harmonics*. The frequencies of the harmonic components are multiples of the fundamental frequency, i.e., $F_n = nF_1$, and are known as *harmonic frequencies*.

The timbre of a sound is determined by the quantities and relative weights of the harmonic components present in the signal. This constitutes what is usually referred to as the *spectral content* of a signal.

1.3.2 Adding up Waves

We can create strange signals by adding simple waves. How strange? In Fig. 1.6, we show a fragment of a periodic signal of a very particular shape, known as a triangular function or *sawtooth*. In the figure, we show how we can approximate the triangular function by superimposing and weighting six harmonic functions. The simulated triangular function becomes more similar to the original function as we keep on adding the right harmonic components to the sum.

A mathematical result widely used in the natural sciences indicates that a large variety of functions of time (for example, that representing the variations of pressure detected by a microphone when we record a note) can be expressed as the sum of simple harmonic functions such as the ones illustrated in Fig. 1.6, with several harmonic frequencies. This means that if the period characterizing our complex note $f(t)$ is T, we can represent it as a sum of harmonic functions of frequencies $F_1 = 1/T$, $F_2 = 2/T$, ..., $F_n = nF_1$, that is, the fundamental frequency and its harmonics:

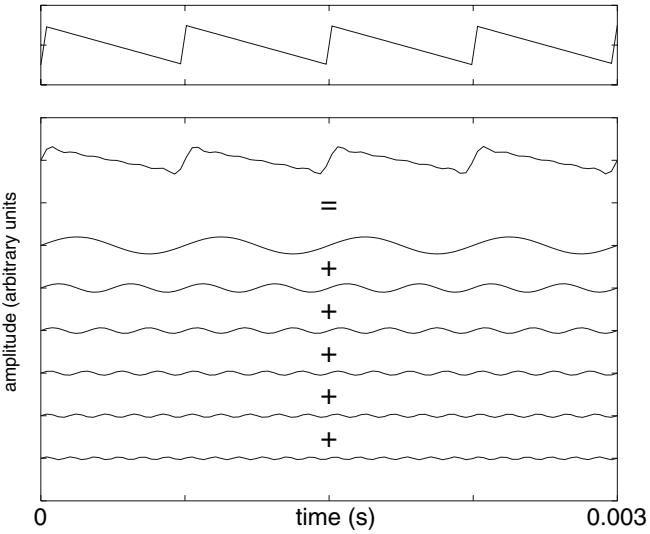

Fig. 1.6. Adding simple waves to create a complex sound. The wave at the top is a complex oscillation known as a triangular or "sawtooth" wave. A simulated sawtooth is shown below, formed by adding the first six harmonic components of the sawtooth. The first component has the same period as the complex sound, the second component has a period half of that (twice the frequency), the third component has one-third the period (three times the frequency), etc. Note that the amplitudes of the harmonic components decrease as we go to higher components. The quantity and relative amplitudes of the harmonic components of a complex sound make up the *spectral content* of the sound. Sounds with different spectral contents are distinguished by our ear: we say they have different *timbres*

$$\begin{aligned} f(t) = a_0 &+ a_1 \cos(\omega_1 t) + b_1 \sin(\omega_1 t) \\ &+ a_2 \cos(2\omega_1 t) + b_2 \sin(2\omega_1 t) \\ &+ \cdots \\ &+ a_n \cos(n\omega_1 t) + b_n \sin(n\omega_1 t) \\ &+ \cdots , \end{aligned} \tag{1.11}$$

where we have used, for notational simplicity, $2\pi F_n = n\omega_1$. Equation (1.11) is known as a *Fourier series*. The specific values of the amplitudes a_n and b_n can be computed by remembering the following equations:

$$\int_0^T \sin(n\omega t)\cos(m\omega t)\,dt = 0\,, \tag{1.12}$$

$$\int_0^T \cos(n\omega t)\cos(m\omega t)\,dt = \begin{cases} 0 & n \neq m\,, \\ T/2 & n = m\,, \end{cases} \tag{1.13}$$

$$\int_0^T \sin(n\omega t)\sin(m\omega t)\,dt = \begin{cases} 0 & n \neq m\,, \\ T/2 & n = m\,. \end{cases} \tag{1.14}$$

The values of the amplitudes are

$$a_0 = \frac{1}{T}\int_0^T f(t)\,dt\,,$$
$$a_n = \frac{2}{T}\int_0^T f(t)\cos(n\omega_1 t)\,dt\,,$$
$$b_n = \frac{2}{T}\int_0^T f(t)\sin(n\omega_1 t)\,dt\,.$$

This set of coefficients constitutes what we call the spectral content of the signal. They prescribe the specific waves for all the harmonic functions that we have to add in order to reconstruct a particular signal $f(t)$. For the moment, it is enough to say that in order to represent a note, we have several elements available: its frequency, its amplitude and its spectral content.

However, there is still some way to go in order to have a useful set of descriptive concepts to study birdsong. If all a bird could produce were simple notes, we would not feel so attracted to the phenomenon. The structure of a song is, typically, a succession of syllables, each one displaying a dynamic structure in terms of frequencies. A syllable can be a sound that rapidly increases its frequency, decreases it, etc. How can we characterize such a dynamic sweep of frequency range?

1.4 Sonograms

1.4.1 Onomatopoeias

Readers of this book have probably had in their hands, at some point, ornithological guides in which a song is described in a more or less onomatopoeic way. Maybe they have also experienced the frustration of noticing, once the song has been identified, that the author's description has little or no similarity to the description that they would have come up with. Can we advance further in the description of a bird's song with the elements that we have described so far? We shall show a way to generate "notes", i.e., a graphical representation of the acoustic features of the song. We shall do so by defining the *sonogram*, a conventional mathematical tool used by researchers in the field, which, with little ambiguity, allows us to describe, read and reproduce a song.

The song of a bird is typically built up from brief vocalizations separated by pauses, which we shall call syllables. In many cases, a bird can produce these vocalizations very rapidly, several per second. In these cases the pauses are so brief that the song appears to be a continuous succession of sounds. But one of the aspects that makes birdsong so rich is that even within a syllable, the bird does not restrict itself to producing a note. On the contrary, each syllable is a sound that, even within its brief duration, displays a rich

temporal evolution in frequency, and is perceived as becoming progressively higher, lower, etc. For this reason, if we were to restrict ourselves to analyzing the spectral content of a syllable as if it was a simple note, we would miss much of the richness of the song. Therefore we use another strategy.

1.4.2 Building a Sonogram

In Fig. 1.7a, we have a signal corresponding to a syllable. We have already worked with it in Fig. 1.2. We shall not look at the complete syllable, but just at a small fraction of it around a given time t. We call this our time window, and we center it around the time t. Let us proceed to analyze the spectral content of this small fragment, and choose the aspects of the spectrum that we find most relevant. We could, for example, concentrate only on the fundamental frequency, forgetting about the harmonics discussed earlier. In this way, we can plot a diagram of fundamental frequency as a function of time, plotting a dot for the fundamental frequency found in the window centered at time t, at that time. For successive times, we proceed in the same way. What we obtain with this procedure is a smooth curve that describes the time evolution of the fundamental frequency within the syllable. This way of analyzing small fragments of a song is a useful procedure for sounds that change rapidly in frequency, and is available as part of almost any computer sound package.

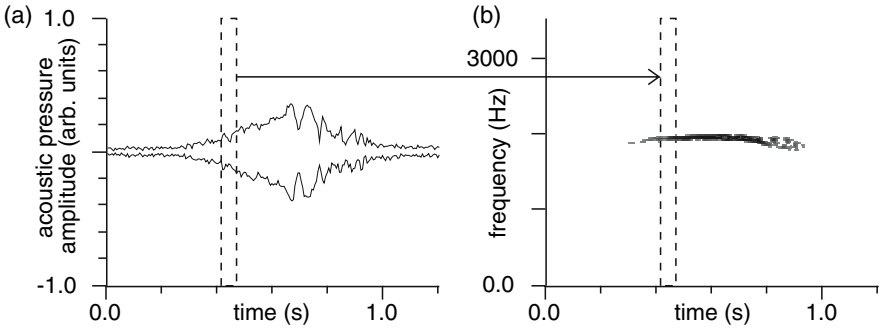

Fig. 1.7. Building a sonogram. (**a**) We start by plotting the sound wave (actually, at this scale, one cannot see the actual oscillation, just the envelope). Now we focus on a very narrow *time window* at the beginning of the recording and calculate the spectral content only for the part of the sound in that narrow window. Next, we slightly shift our time window and repeat the procedure time after time, until we reach the end of the recording. By gathering together all the results we have obtained with the time-windowing procedure, we finally obtain (**b**), the sonogram, which tells us how the sound frequency (and, in general, the spectral content) evolves in time. In this case, the syllable is a note with an almost constant 2 kHz frequency

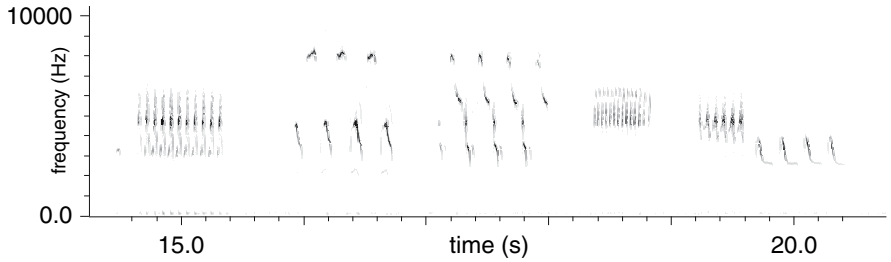

Fig. 1.8. A complex sonogram. This is the sonogram of a grass wren's song. Here you can see *upsweeps* (increasing fundamental frequency), *downsweeps* (decreasing fundamental frequency) and *tonal* sounds (constant frequency). The rich vocalizations of this wren span a huge range in frequency, from around 2000 to 8000 Hz

In Fig. 1.7b, we see the results of the procedure. The time is on the horizontal axis, while the vertical axis indicates the value of the frequency. The fact that the curve is an almost horizontal stroke indicates that the sound, within this brief syllable, has an almost constant frequency – a note. This type of representation is known as a *sonogram*. Of course, instead of focusing only on the fundamental frequency, one can take into account the fundamental and all other frequencies that appear in the spectrum in each time window. The resulting sonogram allows us not only to track the time evolution of the fundamental frequency but also to have a picture of the complete spectrum, evolving in time. In very complex songs, other frequencies appearing in the sonogram may have an evolution different from that of the fundamental frequency.

In Fig. 1.8, we show a sonogram corresponding to the song of the grass wren (*Cistothorus platensis*) [Straneck 1990a]. Notice that the structure of this song is extremely complex: syllables are repeated in rapid succession before being replaced by others that are qualitatively different. Some are upsweeps (the fundamental frequency increases), some are downsweeps, and others are tonal sounds (and therefore could have been described as simple notes). In this book, we intend to provide an understanding to understand what physical processes are at play in generating such a variety of structures. In the process of introducing some elements for the description of these phenomena, we have discussed this graphical representation here, which will allow us to represent the songs that we hear, diminishing the ambiguity of the onomatopoeic description.

2 Sources and Filters

To know that under certain conditions air is capable of propagating pressure (or density) perturbations is the first step in our understanding of sound. However, we still need to discuss a couple of issues in order to continue our presentation on birdsong. On one hand, we are interested in the mechanism used by the bird in order to generate these perturbations. On the other hand, we need to know what happens to these perturbations in the space between where they originate and the open air. Before we analyze the structure of the avian vocal organ, it will be useful to have a general picture of the process of the generation of sounds and the way in which they are filtered. This chapter is dedicated to describing these phenomena.

2.1 Sources of Sound

2.1.1 Flow, Air Density and Pressure

We can begin our discussion by thinking about sources of sound we are familiar with. Examples could include a siren, a flute or the sound that is produced when we blow air between two sheets of paper [Titze 1994]. What do these processes have in common? What physical phenomenon is at play in these cases? As we saw in the previous chapter, sound is a pressure (or density) perturbation that propagates in a medium – in the case we are interested in, the atmosphere. How can we generate such a perturbation?

In order to describe the process by which sound is generated, we must introduce the concept of flow. It is not a complex concept: we constantly refer to traffic flow, the flow of a liquid through a pipe, etc. A flow (of something) is the amount (of that thing) that passes through a surface in certain time. For example, the flow of cars passing through a tollgate on a highway is the number of cars that go through the gate in a certain interval of time. The concept has implicit in it the existence of an area and a velocity. The flow of cars can increase because more cars pass through the gates in the interval of time (i.e., the "velocity" is increased). But let us imagine that the gates are already letting as many cars pass per unit of time as possible: we can also increase the flow by increasing the number of gates (i.e., by increasing

the "area"). In fact, the flow is defined as the product of the area and the velocity of the objects:

$$U = \mathbf{v} \cdot \mathbf{A} \,. \tag{2.1}$$

Here we let the velocity *and* the area be vectors, because the important thing here is the cross section; that is, the effective area faced by the velocity. This can be seen by expressing the dot product as $U = vA\cos\alpha$, where α is the angle between the velocity and the direction perpendicular to the area.

Flow is an appropriate concept for stating conservation laws. For a general closed surface S, we can write the flow U across it as

$$U = \oint_S \mathbf{v} \cdot d\mathbf{a} \,. \tag{2.2}$$

According to our previous discussion, this should be proportional to the variation of the mass within the volume enclosed by the surface S:

$$U = -\frac{1}{\rho_0} \frac{\partial m}{\partial t} \,, \tag{2.3}$$

where ρ_0 stands for the constant equilibrium density of the fluid. Changing from mass to density by means of a volume integral ($\int_V \rho \, dV = m$), we can write

$$U = -\frac{1}{\rho_0} \frac{\partial}{\partial t} \int_V \rho \, dV \tag{2.4}$$

$$= \oint_S \mathbf{v} \cdot d\mathbf{a} \,, \tag{2.5}$$

and since $\oint_S \mathbf{v} \cdot d\mathbf{a} = \int_V \nabla \cdot \mathbf{v} \, dV$ (Gauss's theorem),

$$\nabla \cdot \mathbf{v} = -\frac{1}{\rho_0} \frac{\partial \rho}{\partial t} \,, \tag{2.6}$$

where $\mathbf{v}$ stands for the *particle velocity*. As is known in acoustics, $\mathbf{v}$ is related to the displacement D introduced in Chap. 1 through $\mathbf{v} = \partial \mathbf{D}/\partial t$. The symbol ∇ is a compact notation for the vector spatial derivative. This equation does not provide us with more physics than its one-dimensional version (1.2). However, it is more general and will allow us to advance beyond a particular geometry. Similarly, we can write Newton's second law in a more general way than in (1.3):

$$\nabla p = -\rho_0 \frac{\partial \mathbf{v}}{\partial t} \,. \tag{2.7}$$

As in the one-dimensional case, we can relate the density fluctuations to the pressure perturbations by means of the linearized equation of state (1.4). Simple algebra then allows us to derive the acoustic-pressure wave equation

$$\frac{\partial^2 p}{\partial t^2} = c^2 \nabla^2 p\,, \tag{2.8}$$

which is the three-dimensional version of (1.7).

Solving this equation for given initial and boundary conditions is not trivial. However, two highly symmetric cases have been thoroughly discussed: the plane wave and the spherical wave. These two read

$$p(\mathbf{x}, t) = A e^{i(\mathbf{k}\cdot\mathbf{x} - wt)} \qquad \text{plane}\,, \tag{2.9}$$

$$p(\mathbf{x}, t) = \frac{A}{r} e^{i(kr - wt)} \qquad \text{spherical}\,, \tag{2.10}$$

where r is the modulus of the radius vector $\mathbf{x}$. In these equations, we write the pressure in terms of complex numbers. The physical quantities we are interested in are the real parts of these expressions. The rationale behind this trick is that taking time or spatial derivatives of harmonic functions in this formalism is as simple as multiplying or dividing by a (complex) number. Let us see this at work. In either geometry, it is possible to derive a relationship between the particle velocity at any point and the pressure fluctuations there, by means of (2.6) or (2.7). Since these equations are linear, we can be sure that a solution with both p and $\mathbf{v}$ oscillating with the same frequency is possible. However, a phase difference between them may appear that depends on the symmetry of the solution, which in turn depends on the geometry of the sources and boundaries. For example, in the planar case, where the spatial derivative is everywhere equivalent to a multiplication by a complex number ik, and the time derivative is equivalent to a multiplication by $-i\omega$, the pressure and the velocity are in phase because the coefficient relating them turns out to be real. However, this is not the case for a spherical wave. Relating p and $\mathbf{v}$ by means of (2.7), we now obtain

$$\rho_0 i\omega \mathbf{v} = \left(-\frac{1}{r} + ik\right) p \mathbf{e}_\mathbf{r}\,, \tag{2.11}$$

which means that the pressure and velocity are no longer in phase. Mathematically, this is reflected in the fact that the coefficient relating them is a complex number $\alpha = (-1/r + ik)/(i\omega\rho_0)$.

Why are we saying that a complex coefficient relating p and $\mathbf{v}$ represents a phase difference between them? First, a real coefficient means that p and $\mathbf{v}$ have zero phase difference, since the value of p at a given instant is that of $\mathbf{v}$ (up to a scale factor). On the other hand, the effect of a purely imaginary coefficient relating p and $\mathbf{v}$ can be seen as the fact that the value of p depends on the value of the *time derivative of* $\mathbf{v}$ (scaled by the imaginary coefficient), which is $\pi/2$ out of phase with respect to $\mathbf{v}$. This is a consequence of the following relationship, valid for complex harmonic functions:

$$i\omega \mathbf{v} = \frac{d\mathbf{v}}{dt}\,. \tag{2.12}$$

For these reasons, p and $\mathbf{v}$ are in general out of phase for generic geometries. This can be written as a relationship between the pressure p and both the flow U and its time derivative as follows:

$$p(t) = RU(t) + I\frac{dU(t)}{dt}\,, \tag{2.13}$$

with appropriate coefficients R and I. The terms in the equation above will not contribute equally to the pressure perturbation $p(t)$ in every case. It may happen that in certain circumstances one of them plays a more important role, while the contribution from the other is negligible. We shall come back to this subject in Chap. 6.

2.1.2 Mechanisms for Generating Sound

Let us begin by discussing the way in which a siren works [Titze 1994] – one of the examples mentioned at the beginning of this chapter. This device has a part that blows air, and a rotating disk with a sequence of holes close to the edge, as displayed in Fig. 2.1a. The disk faces the mechanism that blows the air, so that the air jet can pass through one of the holes of the disk. If the disk is set into a rotating motion (for example, with the help of a motor), there will be an airflow through the device that is sequentially established and interrupted as the holes pass in front of the mechanism blowing the air. Let us imagine the process in detail, taking into account the example of the cars discussed in Sect. 2.1.1. One of the holes lets an air jet pass through. In

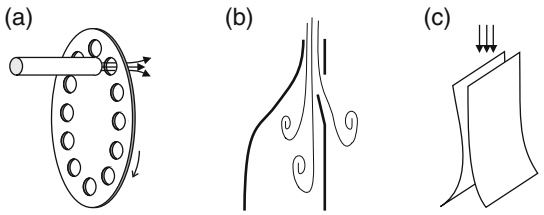

Fig. 2.1. Different physical phenomena generating sound. (**a**) A siren. An air jet is blown against a rotating disk with holes. The air jet passes through a hole only if the hole is just in front of it; otherwise, the jet is interrupted. A pulsating airflow is established in this way. At a constant rotation speed, the siren produces a note (a sound of constant frequency), and the frequency of the note is the frequency at which the holes pass in front of the air jet. (**b**) A recorder. Very close to the mouthpiece, there is a sharp edge that breaks the otherwise smooth, constant airflow into vortices, creating turbulence. The frequency of the resulting sound is related to the effective length of the tube, which can be set by fingering. (**c**) Blowing air between two sheets of paper. Energy is transferred from the airflow to the sheets of paper in such a way that a *self-sustained* oscillation of the sheets takes place. This mechanism is very similar to the one that causes the oscillations of the labia in a bird's syrinx or the vocal folds in a human larynx

the neighborhood of the hole, we have a high density of air. This injection of air will stop as soon as the disk rotates a little, such that the hole no longer faces the mechanism blowing the air. If the propagation speed of the sound is high with respect to the speed of the holes, by the time new air injection takes place, the air density in the neighborhood of the hole will be similar to the density that we had at the beginning of the process. The recurrent passing of holes in front of the air jet as the disk rotates then generates a note. The frequency of such a note will be the frequency at which the holes pass in front of the air jet. For example, if the disk rotates at such a speed that 261 holes pass in front of the air jet per second, then the frequency of the note is 261 Hz (the middle C).

The flute works in a different way. It consists of an open tube made of wood or metal, which becomes narrow close to one end: the end through which we blow air. As the air goes into the tube through this end, it encounters a sharp edge (Fig. 2.1b). The role of this edge during the onset of the sound generation process is to divide the airflow and generate turbulence: the laminar flow loses stability, giving rise to vortices (known as *eolian noise*) that travel downstream along the tube. After this initial process, the jet will alternately leave the edge above or below, owing to a mechanism that involves waves being generated in the tube, giving rise to density fluctuations of a definite frequency. In the siren, a time-varying flow is generated by a mechanical process, while in the flute, a time-varying flow is present because the constant, laminar flow loses stability with respect to a time-fluctuating regime.

It is worthwhile to think about a third example: blowing air between two sheets of paper (Fig. 2.1c). If the sheets are not too large (for instance, one-quarter of a letter- or A4-sized sheet), a good vibrating effect can be achieved. A similar way to obtain a good sound, however, is to cut a piece of paper and hold it between the fingers, as shown in Fig. 2.2. As we blow, we can feel in our lips the paper vibrating, and hear a high-pitched sound. The mechanism by which the sound is produced is not trivial. In fact, it shares many elements with the process involved in the generation of sound in the avian vocal organ. For this reason, we shall analyze this process later. For now it is enough to say that, as in the example of the siren, the sound is produced by temporal fluctuations in the airflow due to periodic obstructions. However, in one important aspect, the physics of this problem is more complicated than that involved in the siren. In the case of the sheets of paper, the periodic obstructions to the airflow are not produced by an "external" motion (such as the rotation of a disk). Instead, energy is transferred from the airflow to the sheets of paper, establishing an oscillatory regime. These oscillations, partially and in a periodic fashion, obstruct the flow. The origin of the sound produced is the creation of local density fluctuations that originate in the presence of a time-varying airflow.

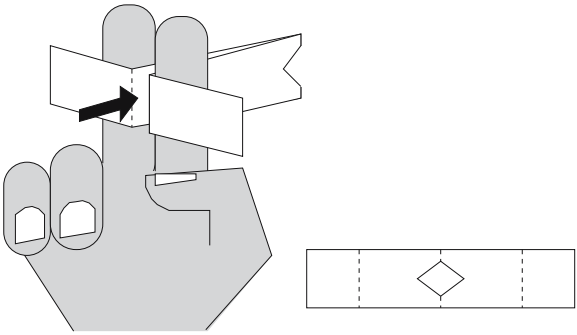

Fig. 2.2. A device with efficient energy transfer from an airflow to moving parts so as to produce a self-sustained oscillation at an audible frequency: in other words, a paper whistle. The physics behind this device is similar to that of the oscillation of the human vocal cords and of the avian labia, and is like that of the paper sheets in Fig. 2.1c. In contrast, an ordinary whistle emits sound in much the same way as a flute (Fig. 2.1b)

2.2 Filters and Resonances

2.2.1 Same Source, Different Sounds

We have discussed the fact that a perturbation in the air density can give rise to a propagative phenomenon. The displacements induced at a point in an air mass as a perturbation arrives are capable of performing work (such as moving the membrane of a microphone, for example). We have also discussed how to generate the original perturbations that will eventually propagate: one way is to establish a time-varying airflow. We still need to discuss another element of importance: the role of passive filters.

In many cases the sound, after being generated, and before propagating freely in the atmosphere, passes through a bounded region of space. The geometry and other characteristics of that region will impose a signature on the sound finally emitted. An example close to us all is the case of the human voice. If you put your hand on your neck, close to the larynx, while you pronounce a vowel, you will feel a vibration. The vibration is produced by oscillations of the vocal folds, which in turn are induced by an airflow from the lungs. This is a sound source whose dynamical nature is very similar to that of the example of the parallel pieces of paper that oscillate when air is blown between them. However, in the case of the vowels, we know that the sound can change dramatically according to where we place our tongue as we vocalize, even if we do not change the oscillation conditions of the vocal folds. In fact, these sounds can be so different that we make them different "vowels". They are just sounds produced with the same source (the vocal cords), but with the configuration of the *filter* changed by moving the tongue, lips, etc. [Titze 1994].

In the first chapter, we discussed the concept of the spectral content of a signal. In the case of the human voice, for example, the time variations of the airflow induced by periodic obstructions caused by the vocal folds typically give rise to signals that are spectrally very rich. By this we mean that they can be written as a sum of many harmonics, as in the case of the triangular wave illustrated in Chap. 1. Our vocal tract stresses some components and attenuates others, modifying the timbre of the sound. How does this happen?

2.2.2 Traveling Waves

Once a source of pressure perturbations has been set oscillating, a propagative phenomenon takes place, leading to a sound wave. As mentioned in Chap. 1, sound waves are solutions to the wave equation (1.7). That is, a sound wave can be mathematically described by a function of space and time $p = p(x, t)$ which fulfills (1.7). One way of achieving this is by choosing p to be a *traveling wave*. Traveling waves are functions of the form

$$p(x, t) = p(kx - \omega t)\,, \tag{2.14}$$

where k and ω are wave parameters such that $c = \omega/k$ is the velocity at which the perturbation propagates, i.e., the sound velocity (see Chap. 1). A well-known example of a traveling wave is the cosine function $p(x, t) = A\cos(kx - \omega t)$. Notice that the particular combination of time and space $kx - \omega t$ in the argument of p is what makes this wave a traveling wave (see successive snapshots in Fig. 2.3a). These waves describe all kind of propagative

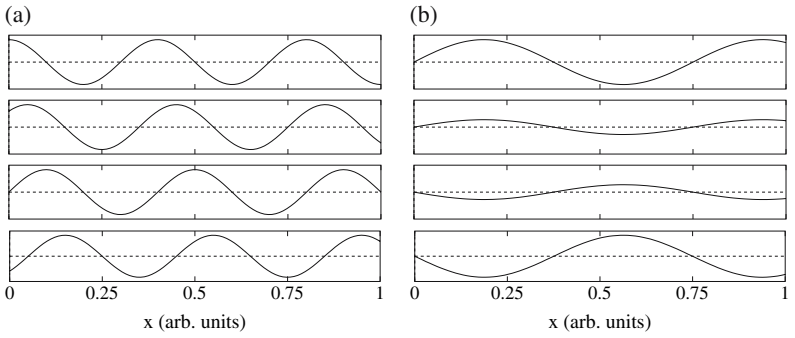

Fig. 2.3. Successive snapshots of waves. (**a**) Traveling wave. A traveling wave is a function $p_T(x, t)$ satisfying (1.7), with space and time appearing in a particular combination: $p_T(x, t) = p_T(kx - \omega t)$. Notice that there are no points at rest, and, further, that the wave is traveling to the right. A wave traveling to the left would be $p_T(x, t) = p_T(kx + \omega t)$. (**b**) Standing wave. A standing wave is a function $p_S(x, t)$ satisfying (1.7) with a factorized space and time dependence $p_S(x, t) = p_1(x)p_2(t)$. Notice the existence of points at rest, or *nodes*: points that always have zero amplitude, being at positions x such that $p_1(x) = 0$

phenomena, from electromagnetic waves [Feynman et al. 1970] to waves of calcium concentration inside a living cell [Keener and Sneyd 1998].

Traveling waves behave interestingly when they reach "boundaries". Imagine a uniform tube, open at one end and closed at the other by a membrane capable of vibrating. The motion of this membrane will push the air in its surroundings, inducing fluctuations of density and pressure that can propagate along the tube. This propagation can be described by a function $p_e = p_e(kx - \omega t)$, as discussed above. Now, what happens as the perturbation of pressure arrives at the open end of the tube? This is very interesting: an important fraction of this pressure perturbation is reflected, *inverted*, so that the sum of the incident pressure perturbation p_e and the perturbation induced by the reflected wave p_r give a negligible total pressure perturbation at the open end of the tube: $p_e(x = L, t) + p_r(x = L, t) \sim 0$. Why? Because the atmosphere imposes its own pressure on the daring attempt of the tube to try to change the pressure with its tiny pressure fluctuations. This is a mechanism similar to that observed when we generate a wave on a string with one end attached to a wall (see Fig. 2.4). The wave propagates and is reflected, inverted. The sum of the displacements (incident and reflected) at the fixed end must be zero, for the end of the string is rigidly attached to the wall. In the case of the tube, the pressure at the open end is "tied" to the atmospheric pressure. Notice that the fact that the pressure fluctuation is zero does not mean that the interior of the tube is isolated from the exterior. The displacement D, as seen in Chap. 1, satisfies $\partial D/\partial x \sim -p$ and therefore it oscillates with the maximum possible amplitude at the open end. These fluctuations of the displacement are responsible for sound emission from the end of the tube.

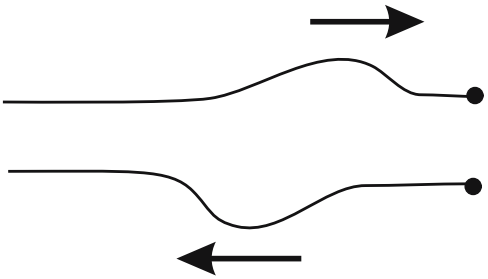

Fig. 2.4. A wave on a string attached to a wall. If you flap the string to create a propagating wave, the wave travels down the string until it is reflected at the end attached to the wall, and becomes inverted. A pressure sound wave propagating along a tube will be reflected and inverted at an *open* end of the tube in an analogous way

2.2.3 Resonances

Now, suppose that our vibrating membrane generates harmonic fluctuations in a periodic way, with a period T. By this we mean that the membrane generates high and low densities in its surroundings (position $x \sim 0$) alternately, in a regular way ($P_e = P_{e0}\cos(\omega t)$, with $\omega = 2\pi/T$), and these fluctuations propagate along the tube. The other end of the tube reflects these fluctuations, which return to the neighborhood of the membrane after traveling back from the open end. Therefore, at every point x along the tube, the pressure perturbation is a superposition of forward- and backward-traveling waves:

$$P_e(x,t) = P_{e0}\cos(kx - \omega t) + P_r\cos(kx + \omega t)\,. \tag{2.15}$$

What happens if, when the membrane is generating a high-pressure fluctuation in its neighborhood, a reflected negative-pressure perturbation arrives? The signal close to the membrane will be the sum of the two contributions, and therefore will be strongly damped. This is known as *destructive interference*. In contrast, if the pressure perturbation, after traveling to the open end and returning back, arrives in phase with the perturbation being generated by the membrane (for example, a high-pressure perturbation arrives as the membrane is compressing its surroundings), the superposition of the signals will be *constructive*. This helps to establish a signal of large amplitude. This phenomenon is called *resonance* [Feynman et al. 1970].

The key quantities for establishing this strong signal in the neighborhood of the membrane are the characteristic times of the problem. In order to construct such a signal, the action of the membrane must be helped by the reflected wave. For this to occur, there must exist a particular relationship between the period of the signal generated by the membrane and the time it takes the signal to propagate to the open end and back after its reflection. When the signal returns, its shape is approximately equal to the shape that it had when it was created, at a time 2τ before (with τ being the time it takes for the sound to travel a distance equal to the length of the tube L). The reflection changes the sign of the signal that returns. Therefore, this perturbation has to be as similar as possible, considering the change of sign that is produced in the reflection at the open end of the tube and the state of the wave on its return, to the signal being created by the membrane, in order to help constructively in the creation of a strong total signal. If the time taken to travel from the membrane to the open end and back is 2τ, a constructive effect will occur if $2\tau = T/2$. The reason is the following: two harmonic oscillations whose phases differ by half a period will be in counterphase, as illustrated in Fig. 2.5. In addition, the reflection of the wave produces an inversion. Therefore, the delay of half a period accumulated in the trip plus the inversion of the wave at its reflection implies that the wave that returns to the membrane after reflection will be in phase with the signal being generated by the membrane. Since $\tau = L/c$ (where c is the sound velocity), we have

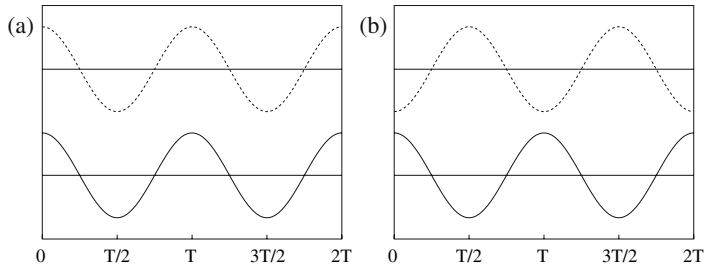

Fig. 2.5. (**a**) Harmonic oscillations *in phase*. Two harmonic oscillations are said to be in phase when both oscillations reach a maximum (or a minimum) at the same time. (**b**) In contrast, they are said to be *in counterphase* when one reaches a maximum when the other reaches a minimum. Notice that two harmonic oscillations in counterphase can be seen as two signals either delayed half a period relative to each other or differing only by an overall factor of -1 (or *inversion*)

a possible frequency of membrane vibrations that will allow the generated wave to be well supported by the tube:

$$F_1 = \frac{c}{4L}\,. \tag{2.16}$$

2.2.4 Modes and Natural Frequencies

The idea is not complex: in order to have a constructive effect, the harmonic signal generated by the membrane must be in phase with the signal that returns after being reflected at the open end of the tube. The theory of waves expresses this idea in the following way: if the tube has a length L and the sound propagates at a speed c, then the membrane must vibrate with a frequency $F_1 = c/(4L)$ in order to contribute constructively and generate a signal with augmented amplitude. This frequency is called the *natural frequency* of the tube, and a pressure fluctuation oscillating at this frequency is called a *mode* of the tube. Since frequency and wavelength are related through $f = c/\lambda$, we have, for this mode, the result that the wavelength λ cannot take any value but $\lambda_1 = 4L$.

If it vibrates at a frequency different from the natural frequency, the membrane can still induce pressure perturbations in the tube, but of a smaller amplitude owing to the reflected wave not arriving exactly in phase with the oscillation generated by the membrane. Is there only one natural frequency for the tube? The answer is no, and this can be understood in the following way. Let us suppose that the membrane is oscillating at just the natural frequency of the tube. As we have discussed, the reflected wave returns in phase with the perturbation established by the membrane. Let us now slightly increase the vibration frequency of the membrane. Now, when the reflected perturbation returns, it will no longer be in phase with the perturbation established

by the membrane. Since the frequency of the vibration is higher now, the perturbations created by the membrane will be "advanced" with respect to the returning wave. Neither the travel time nor the sign inversion at the end changes for the reflected wave. Since the two signals in the neighborhood of the membrane are no longer in phase, the total amplitude of the signal will not be as large as before (when the membrane was oscillating at the natural frequency of the tube).

If we increase the membrane frequency even more, the phase difference between the perturbations created by the membrane and the ones induced by the reflected wave will be larger. Eventually, there will be a frequency for which the two contributions to the total fluctuation will be completely out of phase. For example, the membrane might be compressing the air in its surroundings at the same time as an expansion has traveled back after reflection at the open end. The interesting aspect of this is that if we increase the frequency even more, the situation will reverse and, eventually, the two contributions to the total pressure fluctuation will be back in phase. The oscillation of the membrane is in this case too fast to use the reflected wave to increase the total fluctuation in its surroundings, but it is now sufficiently fast to take advantage of its "second chance" to increase the total amount of fluctuation. The argument can be repeated, and in principle we can see that there are an "infinite" number of natural frequencies.

The concept of natural frequencies is illustrated in Fig. 2.6, where we show the gain of a tube that is open at one end and closed at the other, as a function of the frequency of excitation. The gain is a function that shows us the response of a filter at every excitation frequency. In our case, this frequency is the one at which we make the membrane oscillate. The peaks of the function are the resonances of the tube: each frequency that corresponds to a peak is a natural frequency. Any time we make the membrane vibrate at any of these frequencies, the air column in the tube will vibrate with increased amplitude. In contrast, if the frequency is anywhere between resonances then the oscillation amplitude of the air column will be zero.

The natural frequencies of the tube are easy to identify if we look at the gain curve, but we have to note that the peaks are not ideally narrow but have a finite width. In any real tube there will be dissipation, i.e., inevitable energy losses. In this case, a wave returning to the exciting membrane will do so with its amplitude greatly reduced owing to energy losses. Therefore neither will the peak at a resonance diverge, nor will the total amplitude of the resulting wave be zero at frequencies different from a natural frequency. In the case without losses, the zero amplitude of the wave at a nonresonant frequency is due to the waves added after succesive bounces not being in phase. The infinite summation of all the positive and negative perturbations will give zero. This cannot happen if losses diminish the amplitudes of the waves after a few bounces.

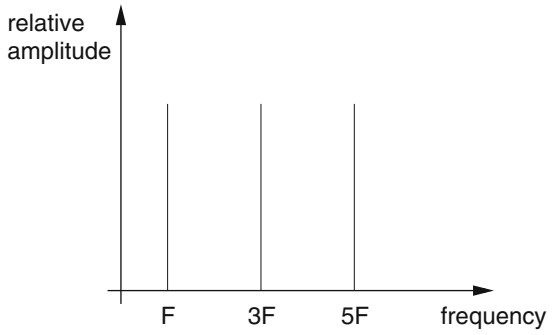

Fig. 2.6. Response function (or gain) for a tube with one end open and the other closed. The response function is the way the tube responds when excited with a sound wave. The peaks are called the *resonances* of the tube. In a real tube, the peaks are not ideally narrow like these but have a width determined by the energy losses. The frequencies corresponding to the maxima of the resonances are the *natural frequencies* of the tube. A sound wave propagating along the tube will *resonate* or increase its amplitude if it has a frequency very close to a natural frequency of the tube. Otherwise, it will be attenuated

2.2.5 Standing Waves

We have discussed the behavior of the density (or pressure) fluctuations in the vicinity of the exciting membrane. It is time to touch upon a most interesting issue, related to the spatial distribution of the density along the tube, when the membrane is vibrating at one of the natural frequencies.

Let us pay attention to the signal generated by the membrane at the first natural frequency $F_1 = c/(4L)$. According to the arguments presented in Sect. 2.2.3, the reflected wave will return to the vicinity of the exciting membrane in such a way that it contributes constructively to the amplitude of the total signal. We say that the reflected wave and the injected signal are in phase, and that the total amplitude of the pressure fluctuations will be maximum. This can be mathematically described as a pattern $p_e = p_e(x,t)$ as follows:

$$p_e = p_{e0}\cos(kx - \omega t) + p_{e0}\cos(kx + \omega t)\,, \tag{2.17}$$

where $k = 2\pi/(4L)$ and $\omega = 2\pi f$. Making use of the trigonometric identity $\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$, we can rewrite (2.17) as

$$p_e = 2p_{e0}\cos(kx)\cos(\omega t)\,, \tag{2.18}$$

which is called a *standing* wave. Functions such as this, with a factorized space and time dependence, are also solutions to the wave equation (1.7) (see Fig. 2.3b for snapshots of a generic standing wave). This result can be interpreted as a temporal oscillation that has a *position-dependent amplitude* $2P_{e0}\cos(kx)$. Notice that at $x = 0$ the amplitude of the oscillation is always

maximum, that at $x = L$ the amplitude of the oscillation is always zero, and that there is no other position at which these situation occur within the tube for this particular choice of ω (and hence k).

What about a signal generated at the second lowest natural frequency of the tube, at which the tube is also capable of sustaining an important excitation? Again, we want the contributions to the density fluctuations in the vicinity of the membrane to be in phase. The wave will contribute (after the perturbation has traveled for a time $2\tau = 2L/c$) constructively to the fluctuations generated by the membrane if the travel time equals a period and a half of the excitation, that is, $3T/2$. Let us recall that in the case of the first resonance, the travel time was equal to half the period of the oscillation. This is so because two signals out of phase by a period and a half will be in counterphase (see Fig. 2.5). Taking into account the inversion that occurs at the reflection, we have the result that the fluctuations induced by the reflected wave are in phase with the ones generated by the membrane. The second natural frequency is therefore

$$F_2 = \frac{3c}{4L}\,, \tag{2.19}$$

which is triple the frequency of the first mode.

Now, something very interesting happens in the tube at a special point. At a distance from the membrane equal to one-third of the length of the tube, the signal that arrives directly from the membrane will be always out of phase by a quarter of an oscillation with respect to the signal at the exciting membrane. The reason is the following: the period of the oscillation is $T_2 = 4L/(3c)$, so $L/(3c)$ (the time it takes to travel a distance $L/3$ at speed c) is $T_2/4$. The signal that arrives at this point in space after undergoing a reflection was emitted some time before: the time interval is the time it takes for the sound to travel 5/3 of the length of the tube. This is so because the wave, before affecting the pressure at our special position in space, had to travel the length of the tube, plus the 2/3 of the tube length from the reflection point to our point of observation. At the frequency of the oscillations considered, the fluctuation induced by the reflected wave has been delayed in this time interval by five-quarters of an oscillation with respect to the fluctuations at the membrane. This is the same as saying that they are one quarter out of phase. This is just the same as the delay of the fluctuation arriving at our observation point directly from the excited membrane! But we still have to consider the change in sign induced in the reflected wave. Therefore the two contributions to the fluctuations, namely that from the wave coming directly from the membrane and that arriving after a reflection, will be exactly out of phase and their contributions will cancel each other. At this particular point, at a distance one-third of the tube length from the membrane, there is no oscillation. The air in this region is exposed to one wave trying to increase the density, and to another one trying to decrease it. This is called a *node*: a

point in space where the amplitude of the fluctuation is zero (in other words, there are no fluctuations).

We can also obtain this result by writing explicitly the spatiotemporal pattern of pressure fluctuations along the tube as we did in (2.18). Adding the forward-traveling wave generated at the vibrating membrane to the reflected, backward-traveling wave, both oscillating at triple the frequency of the first case (i.e., at the second resonance), we obtain

$$\begin{aligned} P_e &= P_{e0}\cos(3kx - 3\omega t) + P_{e0}\cos(3kx + 3\omega t) \\ &= 2P_{e0}\cos(3kx)\cos(3\omega t)\,, \end{aligned} \tag{2.20}$$

which is a standing wave like that of the first mode, (2.18). Notice that now there are *two* points at which the amplitude of the pattern is always zero. One is the open end as before, and the other is, as we expect, at $x = L/3$.

Figure 2.7a shows the first two spatial configurations associated with the excitation of a tube at the two lowest resonant frequencies. In this figure, the lines describe the maximum amplitude of the oscillations that can occur at each point of space. The points at which the lines touch are points with no

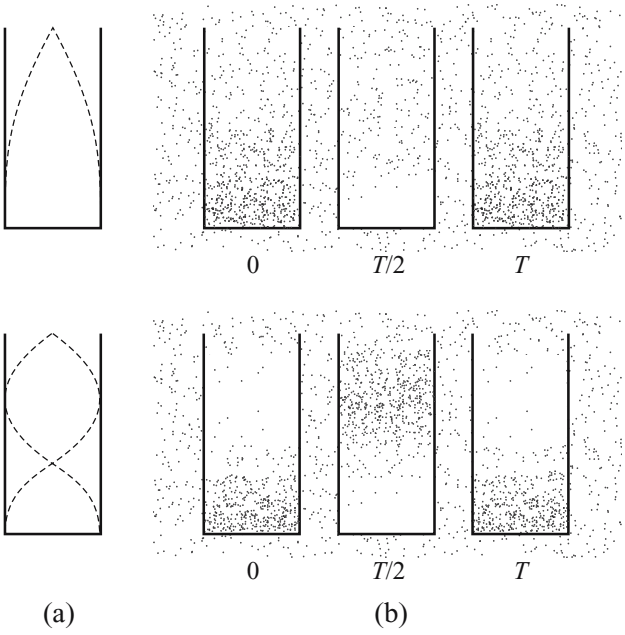

Fig. 2.7. Standing sound waves in a tube with one end open and the other closed. (**a**) Position-dependent amplitude of the oscillation, and (**b**) successive snapshots of the spatial configuration of the air pressure within the tube when the tube is excited at its lowest natural frequency (top) and at its second lowest natural frequency (bottom), the first and second peaks in Fig. 2.6. A higher density of dots means higher pressure

oscillations. In Fig. 2.7b, we show the temporal sequence of the air pressure for each configuration.

As a general result, we can add two counterpropagating waves of the same frequency to obtain

$$\cos(kx - \omega t) + \cos(kx + \omega t) = 2\cos(kx)\cos(\omega t)\,, \tag{2.21}$$

which is a standing wave with a position-dependent amplitude $2\cos(kx)$. The boundary condition at the open end of the tube (that is, the pressure at $x = L$ is always zero, or atmospheric pressure) gives us

$$k_n L = (2n-1)\pi/2\,, \qquad n = 1, 2, 3, \ldots\,. \tag{2.22}$$

This means that the frequency and wavelength of a standing wave in a tube open at one end and closed at the other cannot take any values, but only the allowed values

$$F_n = (2n-1)\,\frac{c}{4L}\,, \tag{2.23}$$

$$\lambda_n = \frac{1}{2n-1}\,4L\,, \tag{2.24}$$

where $n = 1, 2, 3, \ldots$. In contrast, in the case of a traveling wave, the frequency and wavelength have no restriction other than $\omega = ck$.

We have to point out that establishing these stationary density configurations requires a source that excites the tube of air at a precise frequency. What happens if we excite the tube with a signal that is the sum of two periodic signals of different frequencies? Let us analyze this by inspecting a simple example: we take as the excitation a signal that is the sum of the first two resonances,

$$\begin{aligned} P &= A\cos(k_1 x - \omega_1 t) + A\cos(k_1 x + \omega_1 t) \\ &\quad + A\cos(k_2 x - \omega_2 t) + A\cos(k_2 x + \omega_2 t) \\ &= 2A\cos(k_1 x)\cos(\omega_1 t) + 2A\cos(k_2 x)\cos(\omega_2 t)\,, \end{aligned} \tag{2.25}$$

which, although it is the sum of two standing waves with different frequencies (with the same amplitude for simplicity), is not a standing wave. Now, we have a pattern of densities that propagates along the tube, respecting only the condition that there are no perturbations at the open end. In Fig. 2.8, we show a temporal succession of the density along the tube. This density no longer constitutes a stationary structure. There are no points within the tube with no fluctuations. On the contrary, we have a fluctuation traveling along the tube. This can be seen, after some tedious trigonometric algebra, by rewriting (2.25) as

$$\begin{aligned} P &= 4A\cos(kx - \omega t)\cos(\Delta k\, x - \Delta\omega\, t) \\ &\quad + 4A\cos(kx + \omega t)\cos(\Delta k\, x + \Delta\omega\, t)\,, \end{aligned} \tag{2.26}$$

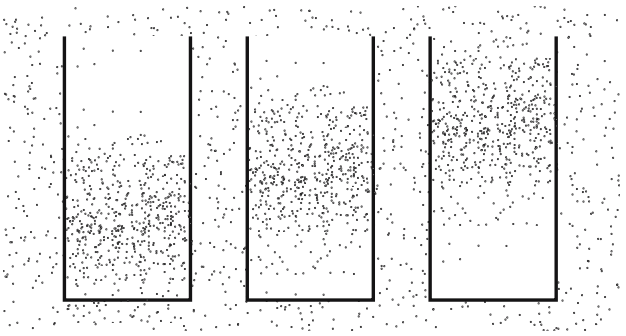

Fig. 2.8. Superposition of two standing waves in a tube open at one end and closed at the other. The resulting wave is not a standing wave, but a traveling wave at the average frequency

where we have defined $k = (k_1 + k_2)/2$, $\Delta k = (k_2 - k_1)/2$, $\omega = (\omega_1 + \omega_2)/2$ and $\Delta\omega = (\omega_2 - \omega_1)/2$. Notice that the resulting wave is a traveling wave (plus its reflection) at the average frequency ω, with an amplitude modulated at a lower frequency $\Delta\omega$.

2.3 Filtering a Signal

2.3.1 Conceptual Filtering

It is time to examine what happens when a signal with a rich spectral content meets a filter. Our discussion of resonances was performed in terms of harmonic excitations produced by a membrane. When did we use the fact that our excitations were harmonic? We assumed this in order to claim that a signal was equal to itself but inverted if we waited for half a period. This is not necessarily true if the signal is not harmonic. Does this mean that for forcings more realistic than simply harmonic ones, we have to forget about our discussion? No, since we saw in Chap. 1 that a periodic signal of arbitrary form can be thought of as a sum of harmonic signals. In this way, we can analyze what happens to a spectrally rich signal in terms of what happens to its harmonic components. Let us take the case, analyzed in detail in Chap. 1, of the triangular time-periodic function. It is spectrally rich (we could approximate it with a sum of properly weighted harmonic functions), and we can describe such richness by showing in a diagram the frequency of each component versus its relative weight. Such a diagram is illustrated in Fig. 2.9. Each peak in the "source" part of the figure represents one of the harmonic components that it is necessary to add in order to reproduce the signal, as discussed in Chap. 1. The frequency at which a peak is located represents the frequency of the component, and its height indicates the relative weight of the component in the sum.

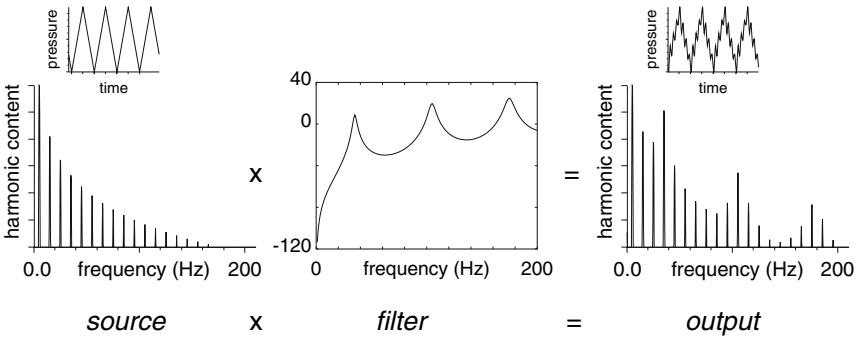

Fig. 2.9. Filtering a signal. *Left*: a source generates a triangular wave (time series at *top left*, spectral content at *bottom left*). This triangular signal is used to excite a tube open at one end and closed at the other (the filter), whose response function is shown in the middle. In accordance with the response function of the tube, some frequencies entering the tube will resonate (the frequencies close to a peak), while others will be attenuated. The result is shown on the *right*: the spectral content of the output signal (*bottom right*) has changed. The time series of the output signal is shown at the *top right*

Now it is time to filter the signal. By this we mean that we excite a tube of air by means of a signal that is no longer harmonic (as in our discussion of resonances and modes), but instead is something like our triangular function. According to our analysis of resonances in a tube, some components of the excitation are capable of establishing important oscillations in the tube, while others will eventually be damped. The gain diagrams discussed in Sect. 2.2.4 gave an idea of precisely those effects. In this way, it is possible to know what will happen to the signal after it has been filtered by the tube: those spectral components falling in the regions of resonant frequencies will survive, while the others will see their relative weight diminished. In Fig. 2.9, we show in addition the gain of a tube open at one end and closed at the other, and the result of filtering our triangular signal. The frequency of the resulting signal is the same as the frequency of the original triangular signal, but the timbre has been changed by the effects of the tube [Titze 1994].

2.3.2 Actual Filtering

There are two ways of actually applying a filter to a signal. The direct way is, obviously, forcing the tube to vibrate by means of a loudspeaker emitting the signal at one end of the tube. This can be accomplished either experimentally (with a real tube) or numerically with the help of a computer, by keeping track of the propagation and reflection of the waves in the tube, which is known as *time-domain filtering*. The details of this procedure are described in Sect. 6.2.2.

The second way is indirect: one needs to know the gain G of the tube as a function of frequency, i.e., the mathematical expression $G = G(\omega)$ giving rise to the central diagram in Fig. 2.9. In addition, the spectral content of the input signal is needed. By this we mean the set of coefficients a_n, b_n in (1.11), represented by a function of (discrete) frequency $P_{in} = P_{in}(\omega_n)$ (recall that a coefficient labeled n stands for the amplitude of frequency ω_n). The spectral content of the signal *after* the filter, P_{out}, is simply

$$P_{out}(\omega_n) = G(\omega_n)P_{in}(\omega_n)\,. \tag{2.27}$$

This is known as *frequency-domain filtering*. Once we know $P_{out}(\omega_n)$, the time-dependent output signal can be easily recovered by summing all the spectral components, as in (1.11).

This is what is called a *passive filter*: the only thing it can do is enhance or attenuate frequencies in the signal's spectrum. It cannot create frequencies that are not present in the original signal. But it should be kept in mind that it is not always possible to separate the effects of the source and the filter. This is a delicate issue. So far, we have assumed that the dynamics of the exciting source are independent of what is happening in the filter, and that the source maintains a strict control of the situation [Laje et al. 2001]. This is known as "source–filter separation". The filter simply adds the injected signal to the successive reflections, selecting some frequencies and suppressing others. However, it is not impossible to think that the signal built up in the tube could affect the dynamics of the exciting source. We shall run into this interesting effect at a later point of our discussion.

2.3.3 The Emission from a Tube

The problem of the tube excited by a membrane at one end and open at the other end is an important exercise in advancing towards our understanding of birdsong: at one end of the tube perturbations are being induced, and the result is the emission of sound at the other end. According to our discussion in the previous sections, the problem reduces to solving the wave equation with the proper boundary conditions. We can propose a solution for the pressure perturbations of the form

$$p = (ae^{ikx} + be^{-ikx})e^{-i\omega t}\,, \tag{2.28}$$

with the boundary conditions

$$v = u_0 e^{-i\omega t} \qquad \text{membrane velocity at } x = 0\,, \tag{2.29}$$

$$p = 0 \qquad \text{at } x = L\,. \tag{2.30}$$

Since the velocity v of the wave satisfies (2.7), we obtain for the amplitudes of the waves

$$ae^{ikL} + be^{-ikL} = 0\,, \tag{2.31}$$

$$\frac{a-b}{c\rho_0} = u_0\,. \tag{2.32}$$

With these coefficients, it is possible to compute the pressure and, therefore, the velocity at the end. The latter has an oscillation amplitude

$$v_{x=L} = \frac{u}{\cos(kL)}\,. \tag{2.33}$$

The divergence at the resonances reflects the idealization of neglecting dissipation. However, it clearly conveys the idea that for some frequencies (those that correspond to $\omega_n = ck_n = c(2n+1)\pi/(2L)$), the particle velocity at the open end will be maximum. Owing to these velocity fluctuations at the open end, there will be variations of the flux that, as discussed at the begining of the chapter, constitute the source of the emitted sound.

3 Anatomy of the Vocal Organ

In Chap. 2, we discussed sources and filters. We stated that the establishment of a time-varying airflow was the basic mechanism by which density fluctuations are created, giving rise to sound. We also described the phenomena due to the existence of a filter between the source and the environment. In that framework, we invited ourselves to touch our neck, at the approximate location of the larynx, while pronouncing a vowel, and we said that as the result of that simple experiment we would perceive a vibration. That perception is due to the motion that is established in a pair of tissues called vocal folds. Why should this discussion be of any interest when we are discussing birdsong? After all, if we were to try to imitate a bird, we would produce some sort of whistle, the physical mechanism of which is certainly well differentiated from the mechanism described above. The whole point of beginning our description by discussing the vibrations of vocal folds in humans during voiced sounds is that, for many species, there are important analogies between this phenomenon and birdsong. And of course differences as well.

3.1 Morphology and Function

3.1.1 General Mechanism of Sound Production

The mechanism of sound production in birds resembles that in humans in that an airflow driven by air sac pressure is modulated by some kind of vibrating valve [Greenwalt 1968]. The acoustic output of this valve excites the air column in the vocal tract. In many cases (both in birds and in humans) this occurs in much the same way as in a source–filter system, and the output of the valve is modified by the resonances of the tract. This mechanism is not the only one to have been proposed to explain the nature of birdsong. A second one, based on an aerodynamic whistle, has been proposed to account for some tonal sounds [Nottebohm 1976, Casey and Gaunt 1985], although direct endoscopic observation [Goller and Larsen 1997b], experiments with light atmospheres [Nowicki 1987] and other studies [Suthers and Zuo 1991] have failed to build confidence in the "whistle" picture.

The mechanism that is emerging as the most adequate one to account for the generation of birdsong can be therefore summarized as one in which

partially adducted membranes are set in motion by Bernoulli forces. This motion affects the airflow, which further affects the dynamics of the membranes modulating the airflow. Song will be produced as long as the motion emerging out of these interactions is of oscillatory nature, since it is capable of modulating the airflow in a way that gives rise to sound.

In agreement with this basic idea, there is a wide diversity of anatomical devices that birds use to produce song. What are the membranes that oscillate? Where are they located? How can the vocal organ be controlled? What are the muscles in charge of this control? In order to describe the physical mechanisms involved in the generation of birdsong, we have to review part of what is known about these issues.

3.1.2 Morphological Diversity

The taxonomic classification of songbirds and nonsongbirds was originally based on syringeal morphology, which varies considerably [Elemans et al. 2003, Suthers 2001]. Most bird species have a *tracheobronchial* syrinx (which is particularly common in the suborder Oscines, order Passeriformes), in which the syringeal membranes are located between tracheal and bronchial cartilaginous rings at the tracheobronchial junction. Two other types of syrinx are found: the *tracheal* syrinx, involving only tracheal cartilaginous rings, occurs mainly in the Furnariidae (ovenbirds) and in some Ciconiidae (storks); and the *bronchial* syrinx, involving only bronchial cartilaginous rings, occurs mainly in Caprimulgiformes (goatsuckers), Cuculiformes (cuckoos) and in some Strigidae (owls) [Casey and Gaunt 1985].

The syringeal anatomy of some birds (such as the ones with tracheal syrinxes mentioned above, and also the parrots, Psittacidae) resembles that of the larynx in humans in that there is only one vocal valve. However, Oscines and other taxa have *two* sound sources, one in each bronchus below its junction with the trachea. Some birds use mostly one source (such as the waterslager canary), but some use both (such as the brown thrasher, *Toxostoma rufum*, with both sides of the syrinx contributing equally to phonation). Some birds can use the two sources independently, producing either the same or a different frequency, and some can even make them interact, as we shall see in Chap. 6. Some syrinxes are displayed in Fig. 3.1.

3.1.3 The Richness of Birdsong

The details of the configuration of the syringeal structure are controlled by the bird by means of a set of muscles, in order to create a wide range of sounds. Remarkably, some muscles at each side of the double structure receive instructions in an independent way. This allows the bird to produce sounds using both sources either in a simultaneous or in a successive way, providing

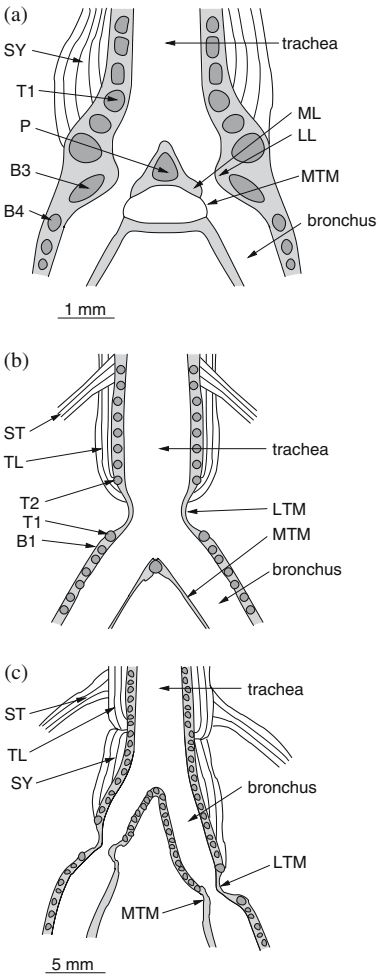

Fig. 3.1. Ventral sections of syrinxes. (**a**) Tracheobronchial syrinx with two sound sources, particularly common in Oscines (brown thrasher, *Toxostoma rufum*). The two vocal valves are formed by adduction of the lateral and medial labia on each side. LL, ML, lateral and medial labia; MTM, medial tympaniform membranes; T1, first tracheal ring; B3, B4, third and fourth bronchial rings; P, ossified pessulus; SY, syringeal muscle. (redrawn from [Larsen and Goller 2002]). (**b**) Tracheal syrinx (ring dove, *Streptopelia risoria*). Constriction of the lumen by the lateral tympaniform membranes forms the only vocal valve. LTM, lateral tympaniform membranes; MTM, medial tympaniform membranes; T1, T2, first and second tracheal rings; B1, first bronchial ring; TL, tracheolateralis muscle; ST, sternotrachealis muscle (redrawn from [Gaunt et al. 1982]). (**c**) Bronchial syrinx (oilbird, *Steatornis caripensis*). Note that the different bronchial lengths, measured from the valves to the trachea, make the syrinx asymmetric. This characteristic has been proved to stamp its signature on the song, which shows two distinct formant bands. LTM, MTM, lateral and medial tympaniform membranes; TL, tracheolateralis muscle; ST, sternotrachealis muscle; SY, syringeal muscle (redrawn from [Suthers 2001])

the sound with possibility of extremely rich spectral features. In Fig. 3.2, we show a sonogram associated with the song of the eastern slaty thrush (*Turdus subalaris*) [Straneck 1990b] in which the two sources are active. Notice that we have a series of notes of approximately constant frequency between 8 kHz and 9 kHz (indicated as "source 1"), and simultaneously we have a second series of notes of a slightly lower frequency (between 7 kHz and 8 kHz, indicated as "source 2") similar to the first, but with independent dynamics.

The richness that the syrinx imprints on the song does not end with the possibility of choosing one of two sources, or singing with two sources at the same time: it is at the level of the syrinx that the bird defines the essential properties of its vocalizations. This is a point of difference between birdsong and the utterance of voiced sounds by humans. The diversity of

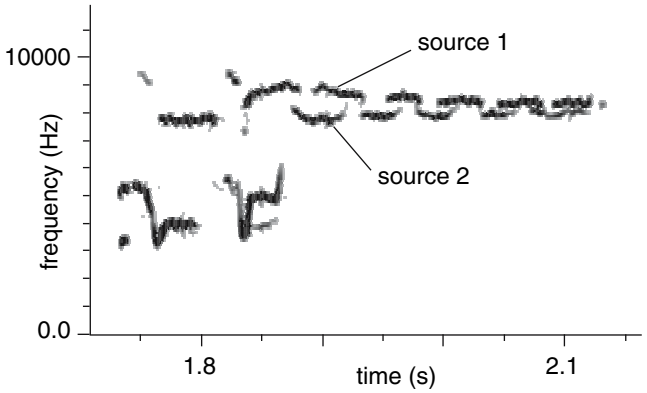

Fig. 3.2. Sonogram of the song of an eastern slaty thrush (*Turdus subalaris*) [Straneck 1990b]. While many birds use only one pair of labia to sing (by keeping the other side silent as a result of active muscular work in order to close it), this bird is singing with *both* sides of the syrinx independently, at the same time. Notice the two sources of sound evolving independently

human voiced sounds is due mainly to reconfigurations of the vocal tract (by tongue position, labial shape, and other configurational measures). The size of the vocal tract, the aperture of the beak, etc., do affect the quality of the vocalization in birdsong. Birds might make such changes in their vocal tract in order to actively coordinate its filter characteristics with the output of the syrinx [Nowicki 1987]. But sound quality and the main properties of the song are primarily defined at the level of the syrinx. In order to understand this point, let us observe the difference between the sonograms in Fig. 3.3.

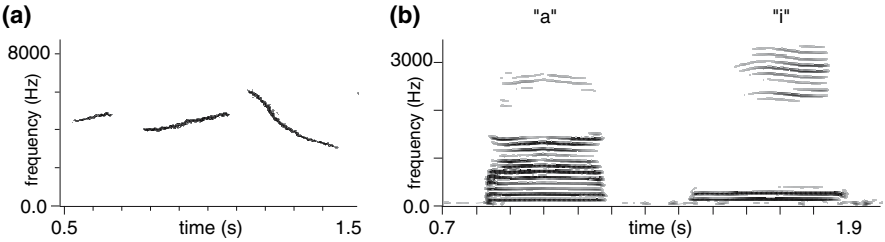

Fig. 3.3. Difference between birdsong and human speech, at the level of spectral and temporal features. (**a**) Three syllables of the chingolo sparrow's song (*Zonotrichia capensis*). Notice the marked time evolution of the fundamental frequency, spanning 3000 Hz in the third syllable. Although it is not evident at this scale, the syllables of the chingolo have almost no harmonic content. (**b**) Sonogram of an utterance of the word "taxi". In contrast to the bird's sonogram shown on the *left*, human vowels have a very rich spectral content and the fundamental frequency (the stroke at the lowest frequency) remains practically constant

The first sonogram shows three consecutive syllables of song of the chingolo sparrow (*Zonotrichia capensis*) [Straneck 1990a], while the second one corresponds to the utterance of the word "taxi" by a human. We notice that, in the second diagram, the fundamental frequency of each vowel remains constant at approximately 120 Hz (this being the frequency at which the vocal folds vibrate). The main difference between the spectra of different vowels is in the quantity and position of the harmonics, which are shown by the traces parallel to the trace of the fundamental frequency. The change in the spectral content, as we discussed in Chap. 2, is due to a change in the morphology of the passive filter (i.e., the position of the tongue in the mouth, the labial configuration, etc.). In contrast, the sonogram associated with the chingolo vocalization shows rich dynamics in the fundamental frequency (moreover, the harmonics are quite weak in this vocalization). The third syllable, for example, begins at 6 kHz and then, displaying an amazing amplitude of register, decreases to 3 kHz.

This shows that it is at the level of the sound sources that the bird gives the song some of its main features. To achieve this, the bird requires a delicate control system that involves the coordinated action of three sets of muscles: those that control the syrinx, those that control the respiratory system, and those in charge of the control of the vocal tract, even if, in principle, they have a less important role [Wild 1993, Wild et al. 1998]. Let us analyze the vocal organ in detail.

3.2 The Oscine Syrinx

Oscine songbirds are capable of vocal learning. This has led to a detailed study of the sophisticated set of neural nuclei responsible for the central control of the song. Oscines also show a high degree of complexity at the level of the vocal organ. For this reason, we dedicate a section to the description of the oscine syrinx.

3.2.1 The Source of Sound

A ventral section of a typical oscine syrinx is displayed in Fig. 3.1a. This illustration is based on a particular species, the brown thrasher (*Toxostoma rufum*) [Goller and Suthers 1996a]. In contrast to the sound-producing organ in humans (the larynx), the syrinx in oscines is a double structure located at the junction between the bronchi and the trachea. All along these pipes we can observe highly modified tracheal and bronchial cartilaginous rings, some surrounded by muscles, and membranes [Goller and Suthers 1996b].

The vocal valves in oscines involve some of the membranes found in the syrinx. These membranes are called the *medial* and *lateral labia* (ML and LL). The labia are membranous connective-tissue masses capable of establishing, under appropriate conditions, oscillations in a way analogous to those

in human vocal folds [Goller and Larsen 1997a]. The valve is formed by constriction of the bronchial lumen when the labia are drawn together by action of adductor muscles. The lateral labium is a thick tissue in the anterior lateral part of the bronchial wall, and it is attached to the third bronchial ring. The medial part of the valve is a membrane attached to half-rings: one side of the membrane is the bronchial lumen, and the other the interbronchial space that opens to the interclavicular air sac. This membrane can be divided into a thick part known as the medial labium and a very thin part called the medial tympaniform membrane (MTM) [Fee 2002]. It is interesting to notice that the MTMs were for a long time assumed to be the sources of sound in birdsong. However, a series of experiments refuted this belief. By painting the MTMs in pigeons with adhesive tissue [Goller and Larsen 1997a] or directly removing them in cardinals and zebra finches, Goller and Larsen established that the membranes are not actually required for vocalization, although the intensity and spectral content of the sound did suffer changes after these alterations.

The identification of the sources of sound is a very important first step in understanding the mechanisms leading to the production of sound. It is then important to understand how this vocal organ is driven in order to generate sounds which can be roughly characterized by their timing, as well as by their spectral properties. This is achieved by the control of a set of muscles that alter the configuration of the vocal organ. The study of the muscular activity during a vocalization is an even harder task than determining the sources of sound. It requires direct electrophysiological measurements of the muscles controlling the syrinx. Such a research program has been performed for several species (for the brown thrasher and the cardinal, see [Goller and Suthers 1996b, Suthers et al. 1999]). The measurements consist in inserting delicate wires that sense the tissue voltage. The voltage measured in this way is a measure of the degree of activity of the muscle. By recording these values for several muscles simultaneously with the airflow and the sound, it was possible to determine the role played by different muscles in the control of the syrinx.

3.2.2 The Role of the Muscles

Figure 3.4 shows another sketch of the syrinx, this time exhibiting the principal muscles attached to it. From the studies described above, it was concluded that some muscles are associated with the active opening or closing of the interlabial space. In other words, the bird can push one labium against the other, actively closing the passage of air. It can also open the passage separating the labia and keeping them apart as far as possible. In both of these situations, the labia will not vibrate. Between these extremal situations, with these muscles relaxed, conditions can be established such that the airflow induces a vibration.

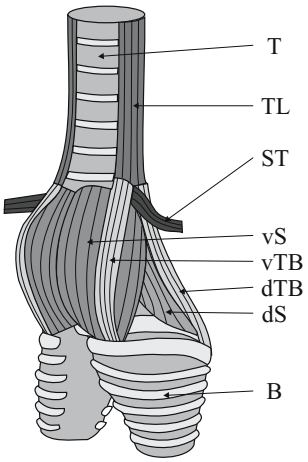

Fig. 3.4. Schematic ventro-lateral view of a songbird syrinx, showing the main muscles involved in the production of birdsong. Muscles (or groups of muscles) have well-defined tasks, for instance controlling the frequency of the vocalization and active gating of the syringeal lumen. Legend: vS, syringealis ventralis muscle; dS, syringealis dorsalis muscle; TL, tracheolateralis muscle; dTB, tracheobronchialis dorsalis muscle; T, trachea; ST m. sternotrachealis; vTB, m. tracheobronchialis ventralis; B, bronchus. Adapted from [Goller and Suthers 1996a]

In studies performed with brown thrashers [Goller and Suthers 1996a], it was observed that the active separation of the labia is controlled by the ventral tracheobronchialis ventralis muscle (vTB) and the tracheoloateralis muscle (TL), while the active closing is mainly controlled by dorsal muscles called the siringealis dorsalis (dS) and the dorsal tracheobronchial (dTB). This was determined by observing that the activity of the first pair of muscles increased significantly while the bird was taking minibreaths (identified by falls of the pressure to values below atmospheric pressure, accompanied by airflow), while the activity of the second pair would increase simultaneously with falls in airflow. In this way, the bird must at least coordinate the muscles that participate in the opening and closing of the syringeal lumen with those involved in respiration, since the pulses of air regulated by the opening and closing of the syrinx are produced by the compression of the posterior air sacs. Note that the bird does not take a deep breath before beginning a song. On the contrary, it takes minibreaths between syllables. This respiratory pattern is replaced by a pulsatile one when a syllable is repeated very fast (more than 30 per second for a canary, for example). In that case, the bird keeps a high level of bronchial pressure (i.e., it does not inspire between syllables) and controls actively the aperture of the syrinx, opening and closing the air passage at the right speed.

On the other hand, in Sect. 3.1 we pointed out that an important difference between a human vocalization and a birdsong syllable (even one

produced by a single source, i.e., by one pair of oscillating labia) was the wide spanning of the vibration frequency. The same set of measurements used to identify the muscles that govern the active opening of the labia also made it possible to recognize the existence of a correlation between the activity of a ventral muscle (siringealis ventralis, vS) and the frequency of the vocalization. The activity of this muscle increases with the frequency. This muscle controls the tension of the labia by means of the separation of the cartilaginous rings that support them.

The role played by these muscles seem to be widespread among oscines [Suthers 2001], although some evidence exists that vS might also play a role in the gating of the sounds produced by zebra finches [Vicario 1991, Goller and Suthers 1996a].

3.2.3 Vocal Learners and Intrinsic Musculature

Songbirds have intrinsic and extrinsic syringeal muscles; the former originate and insert completely within the syrinx, and the latter originate or insert outside the syrinx. Intrinsic muscles probably control the relative position of syringeal elements, while extrinsic muscles affect the syrinx as a whole [Larsen and Goller 2002, Gaunt 1983].

The distinction based on syrinx morphology has survived, although nowadays we are more interested in other aspects of birdsong, namely its relevance as a learned complex behavior. From this viewpoint, independent control of syringeal components seems to be a necessary (although not sufficient) condition for vocal plasticity, since it releases the constraints of a simple syrinx [Gaunt 1983]. Extrinsic muscles are common to all birds, but not all birds have in addition intrinsic muscles which directly alter the syringeal configuration. Indeed, vocal learning seems to be confined to species possessing intrinsic musculature, independent of the number of intrinsic muscles. Among vocal learners, parrots and hummingbirds have two pairs, while some songbirds have at least four pairs, such as the mockingbird (*Mimus polyglottos*).

As has already been mentioned, not every avian syrinx is a bipartite structure. In contrast to songbirds, parrots and doves have only one vocal valve. Direct endoscopic observation of syringeal movements during phonation have shown that oscines and nonoscines also differ in the source of vocalizations: pigeons and parrots use the lateral tympaniform membranes to generate sound, in contrast to the medial and lateral labia used by the oscines [Suthers 2001]. In both cases the sound generator is a vibrating membrane.

3.3 The Nonoscine Syrinx

The syringeal anatomy of the nonoscines presents a wide diversity. In some cases, the syrinx is tracheobronchial as in songbirds. In others, the syrinx is entirely tracheal, and there are cases in which the syrinx is bronchial. In all

cases, the functions of muscles are adapted to allow not only singing, but respiration as well [Suthers 2001]. This can be clearly seen when it comes to terminating a vocalization. This can be achieved either by withdrawing the membranes from the airflow, or by adducting them, blocking the airflow. In birds with a tracheal syrinx, the second possibility is not an option, since adduction would not allow silent respiration. Some cases have been studied in detail. Parrots and pigeons are the most thoroughly studied of the nonoscines. In the following subsection, we shall review the structure of the syrinx of the pigeons. This example can give us an insight into the similarities and differences between oscines and nonoscines.

3.3.1 The Example of the Pigeons

In Fig. 3.1b, we display a schematic ventral view of the pigeon syrinx. A detailed study of the biomechanics of the syrinx in anesthetized pigeons was performed by Larsen and Goller [Goller and Larsen 1997a], who injected gas into the subsyringeal air sacs while various muscles were electrically stimulated. The results of those studies allowed them to identify the sound sources: it was found that air-induced phonation was associated with vibrations of the lateral tympaniform membranes. The control of the sounds in this case has to be carried out with a smaller number of muscles. Contraction of the tracheolateralis muscle (TL) withdraws the lateral tympaniform membranes out of the lumen, opening the air path, while shortening of the sternotrachealis (ST) brings the cartilages of the syrinx closer together, which leads to a folding of the LTMs into the lumen (although even maximal ST contraction does not lead to closure of the syrinx).

In order to start phonation, the pressure in the air sacs is increased. In particular, the interclavicular sac is inflated, which pushes the LTMs into the syringeal lumen, creating a sort of valve. The air passing through these folded membranes sets them into vibration, which modulates, in turn, the airflow going into the trachea. It was observed [Goller and Larsen 1997b] that strong stimulation of the TL muscle leads to termination of the phonation, by the abduction of the membranes from the air pathway. Interestingly enough, weak stimulation of the same muscle was observed to lead to changes in the acoustic features of the phonation. In fact, recent work with doves has shown that TL activity correlates with frequency changes even during fast trills, showing that syringeal muscles have superfast kinetics [Elemans et al. 2004].

Beyond what is known for the pigeons, the birds lacking the complex set of intrinsic muscles that oscines have control the fundamental frequencies of their vocalizations by processes which are not completely understood [Suthers 2001]. It is known that the extrinsic muscles gating the sound affect the tension of the oscillating tissues, as does the subsyringeal pressure, by stretching them [Beckers et al. 2003b]. It is likely that different species have different ways of controlling their vocalizations by combinations of these basic gestures.

3.4 Respiration

The production of song requires precise respiratory patterns, in order to both accomplish adequate gas exchange and provide the airflow needed for phonation. In order to achieve these goals, birds have a respiratory system which is unique in structure and efficiency. The avian lungs are small and relatively rigid (with a nearly fixed volume), and do not move freely [Hildebrand 1995]. Their ventilation is carried out by a set of *air sacs*, thin-walled structures which connect to the trachea and lungs in a complex way. The typical structure consists of a pair of abdominal sacs (where the primary bronchi terminate), a pair of posterior thoracic sacs, a pair of anterior thoracic sacs, paired cervical sacs and an unpaired interclavicular air sac.

The motion of these sacs is driven by inspiratory and expiratory muscles. By moving the sternum ventrally, the inspiratory muscles expand the air sacs, decreasing the pressure. The motion is reversed by the expiratory muscles, which therefore compress the air sacs, increasing the pressure. It is during expiration that most vocalizations are achieved. The pressure during singing can be over an order of magnitude larger than during normal, silent respiration, reaching values of up to 30 cm H_2O in some syllables sung by a canary [Hartley and Suthers 1989].

Despite the wide variety of anatomical features found in the avian vocal organs, it is interesting that some motor patterns are found across different avian groups. For example, in the case of both oscines and nonoscines, the production of repeated syllables at a moderate rate involves a brief inspiration, or minibreath [Calder 1970], which allows the replacement of the air used for the vocalization. Notice that the neural instructions controlling the activity of the inspiratory and expiratory muscles will be interacting with the mechanics of the body during the process. On one hand, stretching receptors will modulate the activity of the neurons controlling inspiration. On the other one, both inspiratory and expiratory muscles will be operating in conjunction with the forces of thoracic elastic recoil: at small air sac volumes, the recoil will help the inspiratory muscles, while at volumes of the air sac larger than the equilibrium value, the elastic recoil will help expiration [Suthers 2004]. A second respiratory pattern frequently used across avian groups is the pulsatile one, used at very high repetition rate. In this pattern, there are no minibreaths, and the expiration is maintained through some level of activity in the expiratory muscles. The critical rate at which a bird will turn to a pulsatile respiratory regime depends on its body size, probably owing to the natural frequency of the mechanical parts of the body involved in the respiration. This hypothesis was tested by tutoring mockingbirds with canaries singing using minibreaths [Zollinger and Suthers 2004]. The result of the experiment was that mockingbirds could copy the high-repetition-rate syllables, but using a gesture that involved pulsatile expiration. This is a beautiful example of the subtle interaction between the nervous system and the body, which we shall explore further in this book.

4 The Sources of Sound in Birdsong

In the previous chapters we discussed the fact that periodic fluctuations in the airflow generate sound, and we have described some mechanisms through which these fluctuations could be established. In this chapter, we shall focus on those mechanisms that are present in the avian vocal organ. The richness of the physics involved in the operating syrinx will lead us to explore several fields, from the physics of fluids to nonlinear dynamics. In order to carry out this discussion, we begin with an analysis of one of the most thoroughly explored and often used models in the natural sciences: the harmonic oscillator.

4.1 Linear Oscillators

4.1.1 A Spring and a Swing

Basically, any periodic motion can be described (at least when the movements are small) in terms of a harmonic oscillator. The harmonic oscillator is a system that performs oscillations as simple as the ones displayed in Fig. 4.1 [Kittel et al. 1965]. In this figure, we show a small mass, attached to a spring, that leaves a record of its motion on a moving sheet of paper. In this example, the motion of the spring is described in terms of the displacement of the mass with respect to its rest position. It is natural to ask about the mechanism used by the spring in order to induce a periodic motion in the mass. The key is that the restitution force that the spring exerts on the mass is proportional to its departure from its rest position (i.e., if the mass is displaced twice as far, then the force applied to the mass by the spring is twice as large). If initially we displace the mass by stretching the spring, the mass will experience a force proportional to the displacement. If we now release the mass, it will be accelerated, increasing its velocity. As the mass passes through the equilibrium position, it will experience no net force (since in this position the spring is stretched by just the amount necessary to compensate the weight of the mass). The inertia of the mass is responsible for the continuation of its motion, and the mass passes through the equilibrium position. At this instant, the mass begins to compress the spring, and a restitution force decelerates the mass until its motion momentarily stops, and it then begins its

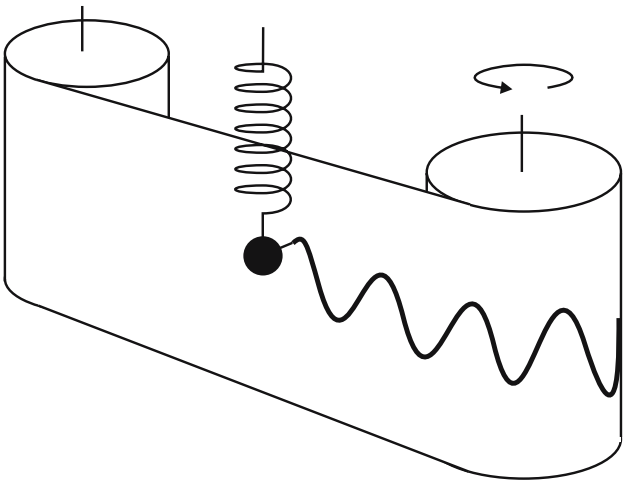

Fig. 4.1. A system performing harmonic oscillation: a small mass attached to a spring. The mass is subjected to a force by the spring, whose characteristic feature is that it is *linearly restitutive*. This means that, if the mass is displaced from its equilibrium position, the spring will tend to restore it to that position by exerting a force proportional to the displacement (that is, the force is twice as large if the mass is displaced twice as far). Once displaced and released, the mass will oscillate back and forth around the equilibrium position owing to the restitution force. The record left on the paper is a trace of the harmonic oscillation

return trip towards the initial position, where the story begins all over again. What determines the frequency of the oscillations of this mass attached to a spring? The strength of the restitution force for a given departure from the equilibrium position. This can be described by means of a restitution coefficient K. If x represents the system's departure from equilibrium, the restitution force is $F_{spring} = -Kx$, and the dynamics of the system can be described by

$$M \frac{d^2x}{dt^2} = -Kx\,, \tag{4.1}$$

or, as a two-dimensional system of first-order equations,

$$\begin{aligned} \dot{x} &= y\,, \\ \dot{y} &= -\frac{K}{M}x\,, \end{aligned} \tag{4.2}$$

with the solution $x(t) = A\cos(\omega t + \phi)$, where $\omega = \sqrt{K/M}$, M is the mass and A and ϕ are determined by the initial conditions. A dot on top of a variable is a shorthand notation for the time derivative.

However, not many of us have seen an oscillating system displaying this behavior for a long time. If we play with a spring, we shall probably see that after a few oscillations, it will stop. Another system which, in principle, can

be described in terms of harmonic oscillations is a pendulum performing small oscillations (like a child on a swing) [Titze 1994]. A child sitting passively on a swing after an initial push will stop after a few oscillations. What force are we forgetting in our discussion?

4.1.2 Energy Losses

We have omitted a force which we can hardly avoid in our daily experience: the friction that occurs during the motion of an object. This is a force that is zero if the mass whose motion we are studying is at rest, but acts on any body that is moving, with an intensity that is proportional to its velocity. The direction of this force is reactive: if the velocity points in a given direction, the force will point in the opposite direction. The constant of proportionality between the friction force and the velocity (let us call it B) expresses how strong the friction is (for example, the friction experienced by a hockey puck sliding on ice is different from what it would be if it was sliding on a wooden floor, although in both cases there is a force proportional (and opposed) to the velocity of the puck). Mathematically speaking, the friction force can be written as

$$F_{friction} = -B\dot{x}\,, \tag{4.3}$$

where $\dot{x} = dx/dt$ is the velocity. In the case of the swing, the friction due to the air interacting with the various oscillating components, and that due to the chains interacting with the structure to which they are attached will be responsible for the loss of the energy initially provided to the system, and will be responsible for the stopping of the motion.

Our understanding of the nature of the agent responsible for stopping our autonomous oscillator gives us a hint of how to counteract the effect of friction. If we want a child on a swing to keep on performing periodic oscillations, we have to push him/her periodically. But this strategy will not work if we push the child as he/she approaches us (exerting on the child a force opposite to the direction of his/her velocity); such a force is precisely what eventually stops the motion. What is appropriate to do is to wait until the child gets to the position of maximal departure from equilibrium, and only push when he/she begins to move away from us. The external force that we are exerting will not cancel the friction at every instant; rather, we are trying to compensate the losses that have taken place in a cycle with the energy provided in each push. If we could push at every instant with a force exactly equal to the friction we could effectively cancel the friction and achieve a perfect harmonic oscillator (amazing everyone around us...). In this way, we exert a force in the same direction as the velocity, that is,

$$F_{gain} = \beta\dot{x}\,. \tag{4.4}$$

If we could exert a force proportional to the velocity at every instant, in the same direction as the velocity, and more intense than the friction, we would

induce larger and larger oscillations; we would be delivering energy to the system.

This description in terms of external forces that compensate the friction is somewhat abstract. We are anxious to understand how these elements could be present in the avian vocal organ, but we still need a few more elements in order to be able to give an adequate description of the mechanisms used to produce birdsong.

4.2 Nonlinear Oscillators

4.2.1 Bounding Motions

In the previous section, we discussed the fact that if we could somehow exert a force proportional to the velocity of a body, we could compensate (and eventually overcome) the effect of the friction force. In terms of the discussion in our previous section, this could be achieved if $\beta > B$. It is particularly pertinent for our problem to understand what happens if the oscillating system has to move within a bounded region of space. At a certain distance from the rest position, for example, there could be walls of some sort. If this is the case, our previous description will remain adequate as long as the body under analysis does not touch the walls. However, if the external force proportional to the velocity that we exert on our oscillating body overcomes the friction, there will be a net force proportional to the velocity of the body, in the same its direction. Therefore the system will oscillate with progressively larger amplitude. Eventually, the system will reach the walls, and a new effect will have to be considered. What happens when the system touches the walls?

The effect of this collision can be modeled in terms of an additional friction, since energy is lost. How can we describe this phenomenon? The usual procedure in physics is to extend what we know works well for a given range of variables to encompass a new range of variables. This enriches both the description and our understanding of the phenomenon under study.

4.2.2 An Additional Dissipation

So far, we have described energy dissipation in a system in terms of friction: a force proportional to the velocity and opposing the system's motion. Now, we want to propose an additional dissipative force that acts only when the body reaches the walls. In this way, the total dissipation is no longer a function of *only* the velocity of the body, but a function of the position as well [Gardner et al. 2001]. We shall not discuss details of this function, which depends on many factors, such as the rigidity of the walls. What is important is that it is a function of the position and therefore the dissipation is no longer a function of only one of the variables of the problem (the velocity). Recall

that the restitution force is a function only of the displacement, and before the walls were introduced, the friction was a function of the velocity only. As we have introduced the walls, the dissipation of the system must be modeled by a force that has a component which depends on both the velocity *and* the position. The problem is no longer linear. This will bring about important changes in the morphology of the oscillatory solutions of the problem.

The simplest conceivable model for our nonlinear dissipative force can be written as $F_{nl} = -Cx^2\dot{x}$. Notice that for small values of x (i.e., close to equilibrium) this term is negligible. It only becomes important when the system is far from equilibrium. Other functional forms that are even in x can be conceived of for F_{nl}, but, since we are not interested in the detailed shape of the solutions for now, we shall retain this form, which is the lowest-order expresion in both x and y consistent with our discussion.

If the force proportional to the velocity F_{gain} overcomes the friction $F_{friction}$, the system will begin to oscillate with progressively larger amplitude, but as soon as the walls are reached, the additional friction force F_{nl} will stop the system. The restitution force F_{spring} will be responsible for the return of the system, which will head towards the opposite wall, to be stopped again. In this way, the oscillations around the equilibrium position will continue, but with a shape quite different from the harmonic oscillations that we would have obtained if all the forces were linear.

Taking into account all these forces, we can use Newton's laws as before to write a differential equation for the midpoint position of a labium:

$$\begin{aligned} \dot{x} &= y\,, \\ \dot{y} &= -kx + (\beta - b)y - cx^2 y\,, \end{aligned} \tag{4.5}$$

whose solutions we shall describe in the following subsection. The constants k, b and c are the already defined K, B and C per unit M.

4.2.3 Nonlinear Forces and Nonlinear Oscillators

In order to illustrate the difference between the predictions of a model of a linear oscillator and the model which we are building, we refer to Fig. 4.2. In this figure, we show the time evolution of the variable x, which measures the departure from equilibrium of the body under analysis. How did we generate this figure? Simply with the help of Newton's equation (4.5), which prescribe the acceleration of a body as a function of the forces acting on it. For given initial values of the position x and the velocity y at some instant t_0, if we know the net force we know the acceleration, and therefore we can calculate the position x and velocity y of the body at an instant $t_0 + \Delta t$ later. In this way, we can progressively build the trajectory of the body. We show the result of this process in Fig. 4.2, for two different scenarios.

In the top part of Fig. 4.2, we show the result for the case in which the body is subjected only to an elastic restitution force, which gives rise to

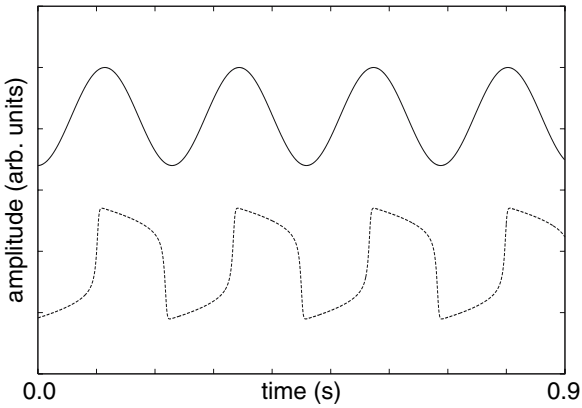

Fig. 4.2. Linear and nonlinear oscillations. (*Top*) time trace of a linear, or harmonic, oscillator such as the one displayed in Fig. 4.1. In terms of the discussion in Chap. 1, this is a simple oscillation. (*Bottom*) time trace of a nonlinear oscillator. The oscillator is a mass subjected to both linear and nonlinear forces. The linear forces are restitution, dissipation and energy supply, and the nonlinear force is a nonlinear dissipation mimicking collisions of the oscillating mass with the wall that it is attached to. This complex oscillation can be decomposed into its harmonic constituents, or components, by means of the representation discussed in Chap. 1

harmonic oscillations. In the bottom part of the figure, we show the time evolution of the body's position when it is subjected to all the forces discussed so far. We have included the linear restitution, the linear dissipation, the external force proportional to the velocity that overcomes the linear dissipation, and the nonlinear dissipation [Arnold et al. 1999], as in (4.5). The nonlinear dissipation is negligible for small values of the departure from equilibrium but large for large departures, which takes account of the collisions of the body with the walls. It is interesting to notice that owing to these collisions, the signal is no longer harmonic. In terms of our discussion in Chap. 1, it is a signal which has several *harmonic components*: it can be thought of as a superposition of several harmonic signals with frequencies that are multiples of a fundamental frequency.

Beyond observation of numerical simulations of the model, we can qualitatively explain the shape of these oscillations. In order to do so, we shall study a dynamical system whose solutions are easy to understand *geometrically*. Then, we shall show how to map our original problem onto this simpler one.

Let us therefore study the following *auxiliary* dynamical system:

$$\dot{u} = v - \frac{1}{3}cu^3 + (\beta - b)u\,, \tag{4.6}$$

$$\dot{v} = -ku\,. \tag{4.7}$$

This system of equations is known as the *van der Pol oscillator*, and it describes the time evolution of two variables u and v. In order to begin our discussion of its solutions, let us notice that two curves in the phase space (u, v) delimit boundaries between regions with different behavior. Within the region of the phase space in which $v > 1/3cu^3 + (\beta - b)u$, u increases, since $\dot{u} > 0$, while u decreases if $v < 1/3cu^3 + (\beta - b)u$. The variable v, on the other hand, decreases whenever $u > 0$, and increases in the region of the phase space where $u < 0$. The curves

$$v = \frac{1}{3}cu^3 + (\beta - b)u\,, \tag{4.8}$$

$$u = 0 \tag{4.9}$$

are known as the *nullclines* of the system. Let us use these curves as a backbone for our understanding of the dynamics displayed by these equations, and let us assume that the constant k in them is small.

We start with an initial condition such that v is larger than $f(u) \equiv -(\beta - b)u + 1/3cu^3$ (see (4.8)). According to (4.6), u increases. A rapid excursion in phase space then occurs until the system reaches the branch with a positive slope of $v = f(u)$ and with positive u (see Fig. 4.3). We say that this excursion is fast, because we compare it with the slow evolution of v: v is small since k is small.

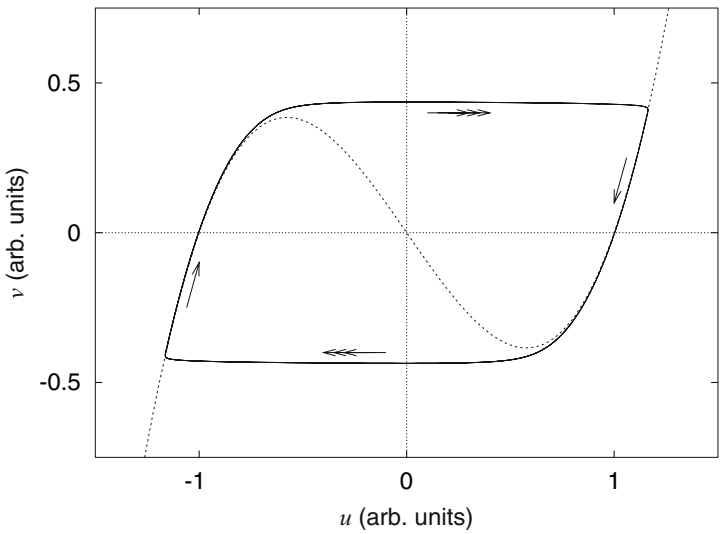

Fig. 4.3. The dynamics of the van der Pol oscillator ((4.6) and (4.7)). The phase space plot illustrates the time evolution of the system (*solid line*), which alternates between slow excursions close to the nullcline $\dot{u} = 0$ (*dashed line*), and fast jumps between branches of the nullcline

Once in the neighborhood of $f(u)$, $\dot{u} \sim 0$ and v slowly decreases, since $\dot{v} = -ku$ and $u > 0$. Notice that the system will evolve so that it "sticks" to the branch of positive slope of the nullcline: if, owing to the decrease in v, the system finds itself below the nullcline, the direction of the force will now be pointing towards the nullcline. The system evolves while trying to keep $v \sim 0$. A qualitative change occurs when the system reaches the minimum of $f(u)$. Now v is smaller than $f(u)$, and therefore u will decrease rapidly. But there is no longer a branch of the nullcline in the region of the phase space with $u > 0$ to stop the dynamics. The system evolves in time in such a way that u decreases, until it reaches the branch with positive slope of $f(u)$ for $u < 0$. Once again, in the neighborhood of $f(u)$, the system will cause v to increase, following $\dot{v} = -ku$ with $u < 0$. Notice that the oscillations that arise alternate rapid jumps with slowly varying time evolutions, as displayed in Fig. 4.2 (bottom trace). These oscillations are known as *relaxation* oscillations.

What we have learned from this system can help us to understand the dynamics of (4.5), since the van der Pol equation can be mapped onto our equations. In order to do so, we have to define $u \equiv x$ and $y \equiv v - cu^3/3 + (\beta - b)u$.

The choice of all the forces mentioned so far in building our model is not arbitrary: we have in mind the processes that will be relevant when we try to understand the operation of the syrinx. And we have discussed almost all the elements that we need.

4.3 Oscillations in the Syrinx

4.3.1 Forces Acting on the Labia

Let us remember that the syrinx is a bipartite device. At each junction between the bronchi and the trachea, there is a pair of labia whose dynamics we want to analyze. It is not difficult to accept that a muscle tissue, after being stretched or compressed, will recover owing to a restitution force. Neither is it difficult to accept that the system has friction (actually, it would be difficult to accept the contrary!). Finally, each labium has a bounded space to move in. When the labia move away from each other, allowing an airflow through, they meet the cartilaginous tubes to which they are attached. Later, within the same oscillation, the labia approach each other until they eventually collide. Consequently, it is also natural to expect a nonlinear dissipation force like the one described earlier. The most difficult problem to be solved in order to establish a description of the dynamics of the syrinx in the framework of our discussion in the last section is to understand the origin of what we have called the "external force". That force is responsible for avoiding the decay of the oscillations due to the dissipation terms.

This problem was addressed for the first time in the study of the dynamics of human vocal folds, and it was Ingo Titze who, in that framework, proposed one of the most satisfactory models for the establishment of membrane oscillations in an airflow. According to this model, the membranes are able to sustain lateral oscillation modes as well as flapping modes. This means that, although the movement of the labia can be very complex as the airflow passes between them, it is possible to decompose this movement into two simpler ones. The first movement consists of the two masses of tissue approaching (or moving away from) each other. In the second movement, the upper edges of the labia move away from (or approach) each other, while the lower edges approach (or move away from, respectively) each other. This second movement can be seen as an upward propagating wave in the mucosa. In Fig. 4.4, we show these basic motions, while Fig. 4.5 displays a series of snapshots corresponding to successive times in the case where both movements are superposed. These motions can be described as follows. Let us call a_1 half the separation between the lower edges of the labia, and a_2 half the separation between the upper edges. If the two modes described above are active, then

$$a_1 = a_{10} + x + \tau y\,, \tag{4.10}$$

$$a_2 = a_{20} + x - \tau y\,, \tag{4.11}$$

where τ is the time it takes for the propagating mucosal wave to travel a distance of half the vertical size of the labia.

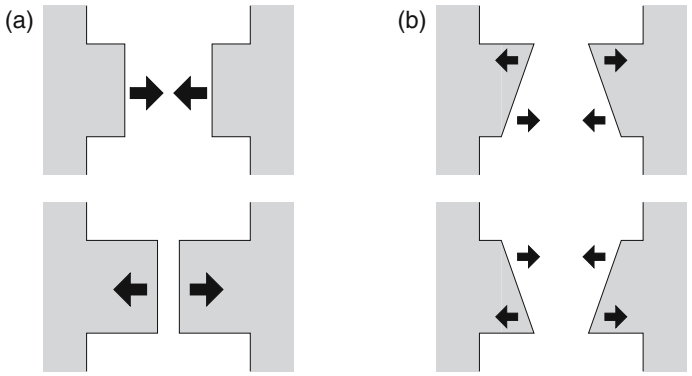

Fig. 4.4. In the flapping model, the labia move according to the coordinated dynamics of two global modes. The first mode is a lateral displacement (**a**), while the second mode is an upward-propagating wave (**b**). If the labia are displaced away from each other and have a convergent profile, or are displaced towards each other and have a divergent profile, the labia gain energy from the airflow in each cycle. The reason is that when they have a convergent profile, the average pressure between the labia is closer to the bronchial pressure, whereas the interlabial pressure is closer to atmospheric pressure for a divergent profile. This results in a force in the same direction as the velocity of displacement of the labia, which can overcome dissipation

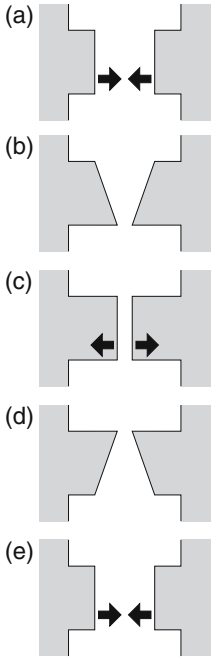

Fig. 4.5. If the two global modes are active in the right phase, the labia gain energy from the airflow. This series of snapshots illustrates the final labial motion in such a situation (modified after [Gardner et al. 2001])

To understand the origin of these movements in an airflow is not trivial, but we are going to suppose that these movements can exist. Moreover, we are going to suppose that they are coordinated in such a way that the labia have a convergent profile while they are moving away from each other, and a divergent profile when they are approaching each other. We shall see that under these assumptions, it is easy to understand the origin of the force responsible for overcoming the dissipation [Titze 1988, Gardner et al. 2001, Laje et al. 2002a]. It is appropriate to point out that this mode structure is compatible with direct endoscopic observations [Goller and Larsen 2002, Fee et al. 1998].

4.3.2 Self-Sustained Oscillations

Let us see now how the combination of movements described above allows the establishment of oscillations. We begin with the bird expiring, meaning that it generates a high bronchial pressure, responsible for the airflow. In this situation, the pressure below the labia is higher than atmospheric pressure, while the pressure above them is essentially equal to atmospheric pressure [Titze 1988, Gardner et al. 2001]. When the labia form a convergent profile (see Fig. 4.6), the value of the pressure between them is nearly the bronchial pressure. On the other hand, when the labia form a divergent profile, the pressure between them is approximately atmospheric pressure. Therefore,

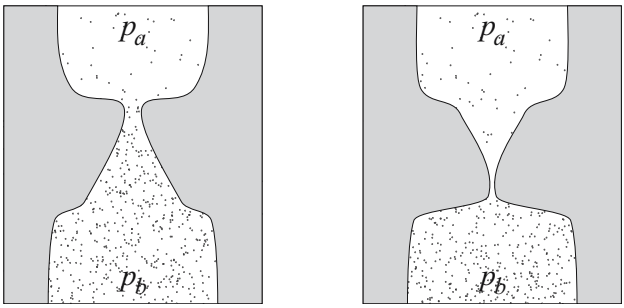

Fig. 4.6. The labia can gain energy from the airflow when the two modes of motion described in the text are active. The reason is that when they have a convergent profile, the average pressure between the labia is closer to the bronchial pressure, whereas the interlabial pressure is closer to atmospheric pressure for a divergent profile. Consequently, if the the labia are displaced away from each other while forming a convergent profile and displaced towards each other while forming a divergent profile, the labia gain energy from the airflow in each cycle. In the first part of the cycle, there is a strong force in the same direction as the velocity. In the second part of the cycle (when the labia move towards each other), the force (which opposes the velocity) is weaker

if the labial movements are coordinated in such a way that they form a convergent profile when they are moving away from each other, there will be a force acting on the labia in the direction of their velocities. Depending on the value of the bronchial pressure p_b, this force may be capable of generating an amount of work that overcomes the energy losses, giving rise to self-sustained oscillations.

Computing the average pressure per unit M (see (4.1)) between the labia gives rise to the following expression:

$$p_f = p_b \left(1 - \frac{a_2}{a_1}\right) , \tag{4.12}$$

where p_b is the sublabial pressure per unit M. The technical details are somewhat involved, but the basic idea is that the coordinated modes are able to sustain oscillations of the membranes, transferring energy from the airflow to the oscillating system [Titze 1988, Laje et al. 2002a]. In fact, Newton's equations now would read

$$\begin{aligned} \dot{x} &= y\,, \\ \dot{y} &= -kx + -by - cx^2 y + p_b \left(1 - \frac{a_2}{a_1}\right) , \end{aligned} \tag{4.13}$$

which, after making the approximation $1 - a_2/a_1 \sim y$, gives rise to the equations of the simple model (4.5). Notice that these equations are similar to those describing a mass attached to a spring, which actually was the first quantitative model of sound production for birdsong [Fletcher 1988].

4.3.3 Controlling the Oscillations

One of the parameters that a bird has to control in order to establish an oscillation in the labia is the bronchial pressure. The bronchial pressure has to be larger than a certain critical value in order to set up the oscillations, which periodically obstruct the airflow. (This periodic obstruction is, as seen in previous chapters, the physical origin of the sound.)

It is important to point out that the labial elasticity k is also controlled by the bird. In the case of the oscine birds, the activity in the siringealis ventralis muscle determines the stretching of the labia (and therefore their stiffness). In Fig. 4.7, we show the spectral content of four different oscillations (obtained from numerical simulations like the ones performed in order to generate Fig. 4.2) for different values of the bronchial pressure p and labial elasticity k.

There are two salient features to note. First, as the labial elasticity k increases, the frequency of the oscillations increases. Second, the bronchial pressure not only determines whether or not there will be oscillations, but also determines the spectral content of the sound. The greater the bronchial pressure, the more violent the collisions between the labia and against the

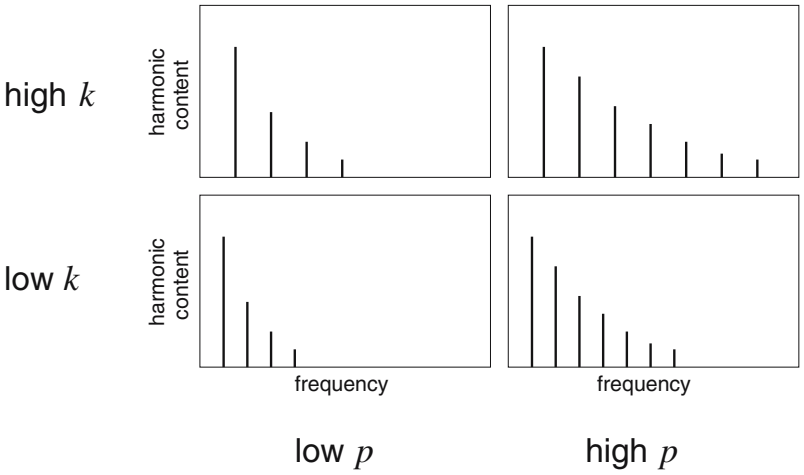

Fig. 4.7. The features of the oscillations that this model generates depend on two parameters: the restitution constant of the labia (k) and the pressure (p). The higher the restitution constant, the higher the fundamental frequency of the oscillations (compare, for the same pressure, the *upper* panels with the *lower* panels). In each of the panels, the spectrum of the resulting oscillations is displayed. On the other hand, as the pressure becomes higher, the oscillations become more "violent". This results in a higher spectral content of the signal, which can be seen from the spectra by comparing how many harmonics are needed to represent the sound. Compare, for the same restitution constant, the panels on the *left* with the panels on the *right*

walls. The increase in the strength of the collisions is responsible for the richer spectral content of the sound produced [Gardner et al. 2001].

There is still one more element to discuss. According to our description of the syrinx in Chap. 3, this is a bipartite device [Goller and Suthers 1995], with a pair of labia in the junction between the bronchi and the trachea. In some species, both sources are used simultaneously, creating rich and complex sounds. However, it is usually the case that one finds a high degree of laterality, i.e., the use of only one source of sound, on the left or right depending on the species [Goller and Suthers 1996a]. In order to achieve this, birds often actively close one of the conduits, forcing the labia on one side to stay pressed against each other. This is done through the activation of dorsal muscles (specifically, the siringealis dorsalis and the tracheobronchialis dorsalis), and in terms of our model of a labium, this corresponds to a force f that does not depend on either x or y. If this term is large enough, the oscillations can be stopped, by forcing the labia to stay pressed against each other.

It is by means of the control of these parameters (control of the bronchial pressure with respiratory muscles, labial tension with ventral muscles and active adduction with dorsal muscles) that the bird is capable of establishing oscillatory dynamics in the labia, periodically obstructing the airflow that feeds the vocal tract [Laje et al. 2002].

The oscine birds have a large set of muscles to control the acoustic features of their vocalizations. Nonoscine birds, however, also have some control [Suthers 2001]. Doves, for example, can alter the tension of the oscillating membranes with the same muscles as used to gate the airflow. On the other hand, the high pressure in the interclavicular air sac inflates it, affecting the tension of the oscillating membranes. These processes contribute in a complex way to determining the acoustic features of nonoscine vocalizations [Beckers et al. 2003b].

4.4 Filtering the Signal

An element lies between the source of sound and what we ultimately hear: the vocal tract, which is the "tube" that runs from the syrinx to the beak. The airflow fluctuations produced in the syrinx (at the entry to the vocal tract) generate, as we have seen in Chap. 2, sound waves that travel through the tract. In regions where the tract alters its shape (for example, by becoming narrower or bending), part of the incident wave will be reflected and part of it will be transmitted. The same happens at the beak. Part travels to the exterior and part of the incoming wave is reflected back into the tract. The consequences of this were discussed in Chap. 2. The basic result is that some spectral components will be reinforced while others will be damped. In Fig. 4.8, we display the song of a pirincho, (*Guira guira*) [Straneck 1990a], in which the presence of a filter is clear in the fluctuations (of half a period)

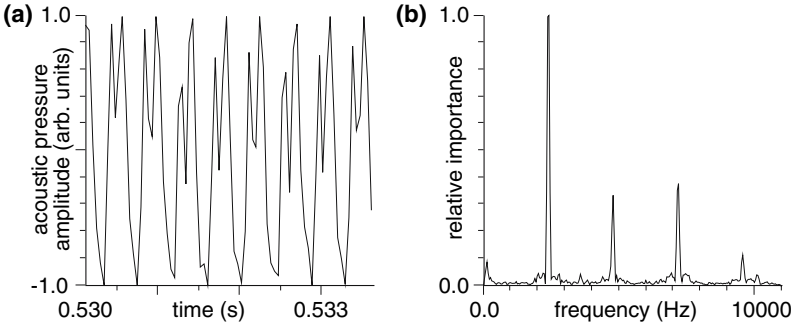

Fig. 4.8. The acoustic pressure (**a**) and Spectrum (**b**) for a note of a *Guira guira*. Notice that the spectrum has many peaks, which do not monotonically decrease as a function of the frequency. This is due to the effect of the filter (passive tract) on the pressure fluctuations induced by labial oscillations

superimposed on the fundamental oscillation at 2333 Hz. It is important to point out that the relative sizes of the harmonics provide the song with its characteristic timbre.

This chapter has dealt with the problem of the generation of sound by means of the induction of oscillations in syringeal labia. The airflow fluctuations induced by these oscillations excite a tube that filters the signal, reinforcing certain frequencies and damping others. We have also discussed qualitatively the influence of the labial elasticity and bronchial pressure on the acoustic features of the sound generated. In the following chapter, we shall advance to the description of the instructions that the bird has to send to its avian vocal organ in order to generate the various sounds that its song is built from. We shall do so in the framework of simple models, similar to those which we have discussed so far.

It is reasonable to keep in mind all the simplifications that we have made. In the model that we have presented, the oscillating labia are symmetric. In our model, the nonlinear dissipation is modeled as a simple polynomial function. We do not take into account the internal structure of the oscillating labia, although the difference in mass between the medial labium and the medial tympaniform membrane might serve to induce the wave-like mode that we are using in our description [Fee 2002]. We have also described the problem of the generation of the sound separately from the problem of the linear filtering of the sound, although the back-propagating waves (from partial reflections at interfaces) could affect the oscillations of the labia, generating extremely rich dynamics. Some of these issues will be revisited in Chap. 6. However, it is important to know how successful the simplest models can be before discussing these subtle effects. The exploration of the simplest models is the focus of Chap. 5.

5 The Instructions for the Syrinx

The diversity of the basic acoustic elements that constitute a bird's song is amazing. In order to create these elemental vocalizations, the brain of the bird has to send a set of precise instructions to the vocal organ. In the case of songbirds, the neural motor pathway involved is composed of four neural nuclei (each with thousands of neurons), which innervates some 20 muscles (syringeal and respiratory). Now that we have some understanding of the mechanics of the avian vocal organ, some questions naturally arise. How does the bird achieve the coordination of so many signals? How complex are the instructions sent by the avian brain to the syrinx in order to create a song?

5.1 The Structure of a Song

5.1.1 Syllables

A bird's song is one of the richest acoustic phenomena found in nature. Typically, it has a complex and varied structure, built out of a series of blocks, each one made from the repetition of brief, continuous vocalizations that we call syllables. In the case of the canary, for example, the syllables last between 15 and 300 ms (i.e., the bird executes several syllables per second). Each bird has several dozen syllables, which are combined in order to create the songs in its repertory. In Fig. 5.1, we show sonograms of two song fragments: one vocalized by an ashy-tailed swift (*Chaetura andrei*) [Straneck 1990b], and the other by a greenish manakin (*Schiffornis virescens*) [Straneck 1990c]. The songs of these two species are globally very different, yet at the level of the syllables, they show important similarities. Although the structure of the harmonics the timbre, the durations or frequency ranges might be different, in both cases the sonograms show a sequence of small continuous "curves". These are typically repeated several times, and are separated by silences. Each continuous curve sweeps a certain frequency range. As we can see, the frequency typically evolves within a syllable, either upwards, or downwards or "*n*" shaped.

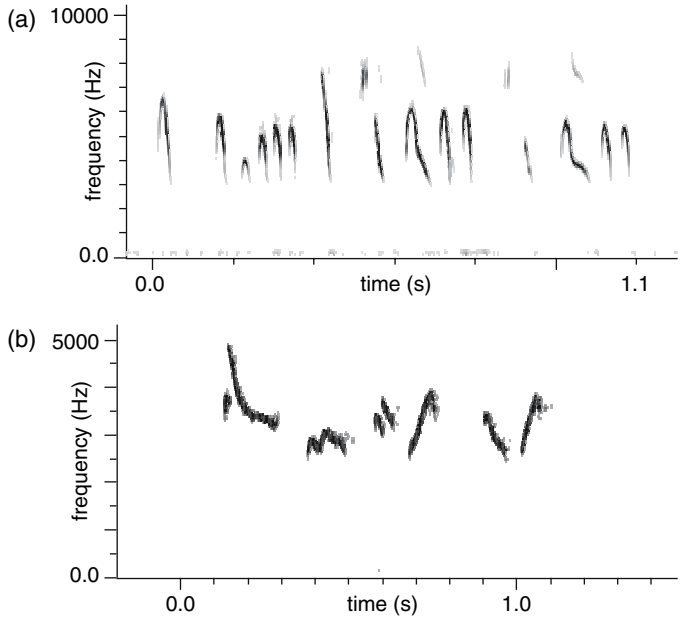

Fig. 5.1. Sonograms of two song fragments: (**a**) one vocalized by an ashy-tailed swift (*Chaetura andrei*), and (**b**) one by a greenish manakin (*Schiffornis virescens*). The songs are globally very different. The timbres of the sounds might differ, as well as the average fundamental frequency of the notes. The combinations of sounds used to compose the song are not the same. However, at the level of the note, there is one important similarity: in both cases the sonogram shows a sequence of small continuous "curves". Each continuous curve sweeps a certain frequency range. The frequency typically evolves within a syllable, either upwards, downwards or "n" shaped

Calder conjectured that the spaces between syllables are used to make the mini-inspirations needed to execute the song [Calder 1970]. This hypothesis was experimentally validated by Hartley [Hartley and Suthers 1989, Hartley 1990]. She found that each syllable is accompanied by an air pulse (expiration), while during each intersyllabic silence, the air sac pressure falls below atmospheric pressure (inspiration) creating an inflow of air. This respiratory strategy is not used in trills repeated at high frequencies, but in canaries, the existence of mini-inspirations was found in executions of more than 30 syllables per second.

The execution of a syllable after a mini-inspiration implies a raising of the air sac pressure. This increase in pressure in the air sac continues until it reaches a value such that, according to our discussion in the previous chapter, the force exerted by the interlabial pressure overcomes the dissipative forces. In this way, the labial oscillations that generate the sound are established. After having sustained these oscillations for a while, the pressure decreases and

the labial motion stops. Now the bird is ready to perform a mini-inspiration before the execution of a new syllable.

5.1.2 Bifurcations

In our description, an increase in pressure has a remarkable consequence for the state of the system. As the pressure grows beyond a critical value, the labia begin to oscillate. This qualitative change in the behavior of a physical system as a parameter is changed is called a *bifurcation* [Solari et al. 1996]. Bifurcation is a concept that refers to any qualitative change in the solutions of a nonlinear problem, as a control parameter is varied. The transition from a stationary state to an oscillatory state is one example. But not the only one.

Some systems show a bifurcation from one static state to another static state. As an example, let us analyze briefly what happens when a pair of forces is exerted on a metal plate (as illustrated in Fig. 5.2a). If the magnitude of the force on the plate is below some critical value, the undeformed plate is *stable*. We can deform the plate a little, and after the perturbation is removed, the system will evolve back towards the undeformed state. The situation is different if the magnitude of the force exceeds the critical value. Then, the

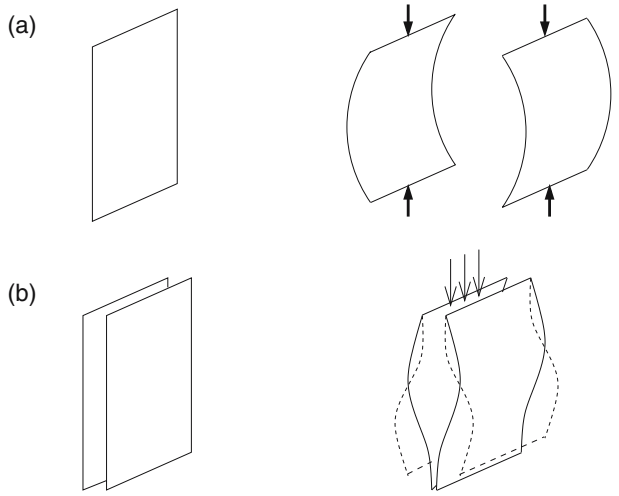

Fig. 5.2. A bifurcation is a qualitative change in the dynamics of a system as a parameter is changed. In this figure, two examples are illustrated. In (**a**), a plate is subjected to a force (hold a credit card between your fingers and try the experiment). If the force is high enough, the plate (card) will bend. The symmetric solution becomes unstable. The initial conditions will determine whether the bending is to the right or to the left, if the plate is perfectly symmetric. In (**b**), we show what happens if you blow air between a pair of paper sheets. If the airflow exceeds a threshold, then the sheets begin to oscillate

system no longer remains in the undeformed state. It is interesting that *the problem* is as symmetric as before (the forces are still on the same axis), but now a slight perturbation in the initial configuration will evolve until the system finds its equilibrium in the bent-plate state. Whether it bends to one side or the other depends on the initial conditions. By the way, the problem is an *extended* one: both space and time are relevant. In fact, we could think of a situation in which the force was so strong that we broke the plate, creating a most complex spatial pattern. But as long as we are interested in discussing only whether the system is in the undeformed state or the bent state, we can restrict ourselves to the study of the dynamics of the *spatial mode*

$$\epsilon = A \sin\left(\frac{\pi}{L}x\right) , \tag{5.1}$$

where ϵ represents the deformation of the plate with respect to the plane where the forces are acting. The situation where the forces are below the critical value has $A = 0$ as a stable solution. Even without knowing the details of the elastic properties in the problem, we can advance a little in our modeling. The time evolution of this mode does not seem to be influenced by any other higher modes of the problem, and therefore we can attempt to model its dynamics in terms of only one variable. The dynamics can therefore be written as

$$\frac{dA}{dt} = f(A) , \tag{5.2}$$

where $f(A)$ depends on all the elastic properties of the plate, as well as the forces. However, we know that since the problem is symmetric, the function f should satisfy the condition that $f(A) = -f(-A)$. An expansion of this function in a power series should therefore be, for small values of A, of the form

$$f(A) = \mu A + cA^3 , \tag{5.3}$$

where μ and c depend on the parameters of the problem. We can advance further in our qualitative modeling. We know that the transition as the external forces are increased can be captured by only the linear terms (since for small values of A, the cubic term is negligible). Therefore, μ should depend on the difference between the external forces and some critical value. Moreover, when $\mu > 0$, we want the nonlinearities to be capable of stopping the growth of A, and therefore c should satisfy the condition that $c < 0$. In this way, we find that for $\mu < 0$, $A = 0$ is stable, and when $\mu > 0$, $A = 0$ is an unstable solution, while the stability is transferred to $A = \sqrt{\mu}$ and $A = -\sqrt{\mu}$. The moral of the tale (or at least one of them) is that even if the problem is an extended one, the qualitative change from a nondeformed state to a deformed one can be described in terms of a low-dimensional system of equations.

Other systems show changes that imply the appearance of motion: for example, the establishment of oscillations in a pair of paper sheets as the airflow between them exceeds a critical value, as shown in Fig. 5.2b. This

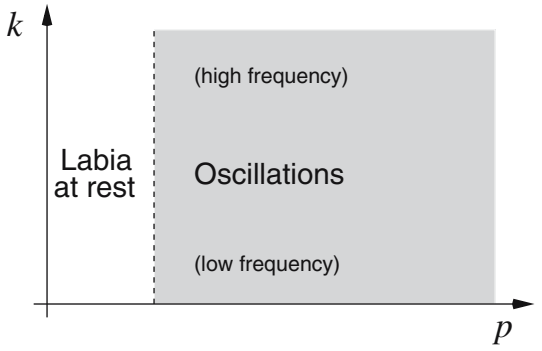

Fig. 5.3. The parameter space of a simple model of sound production. The axes are the restitution constant (K) and the pressure (P). The *shaded* region denotes the region of the parameter space in which oscillations take place. If, initially, the system operates with parameters outside the *shaded* region, the labia will not move. If the parameters are slowly changed towards the *shaded* region, oscillations will start as we enter it (modified after [Laje et al. 2002])

example belongs to the class of problems that we must understand in order to address the issue of labial oscillations. The appearence of oscillations with a well-defined frequency is known as a *Hopf bifurcation*. In the case of labial oscillations, this frequency depends on the value of the restitution coefficient of the labia when the oscillations begin.

In Fig. 5.3, we show the states of the syringeal labia for a range of values of the pressure and muscle tension. The pressure and the tension constitute what we call the parameter space of the system. To the left of the vertical dashed line, the pressure is not large enough to establish oscillations, and therefore the labia remain at rest. In contrast, as the pressure is increased and the vertical line is passed, oscillations are started (shaded region). Depending on the tension as the critical pressure is passed, oscillations of different frequencies are started. The turning off of the oscillations when the pressure is decreased is described in terms of an inverse Hopf bifurcation.

The model described in the previous chapter to describe the dynamics of the midpoint of a labium also assumed a restricted number of active modes, as in the example where we discussed the deformation of a plate. In the case of the labial motion, we assumed a lateral mode and a wavelike mode in order to allow the transfer of energy from the airflow to the labia. By assuming a small number of modes, we restrict the level of complexity that the solutions of the model can display. We have to add this assumption to our list, in order to explore the possible consequences of removing it. As with other effects, we shall deal with this question in Chap. 6. Now, we shall continue with our description of the simple, *low-dimensional* model presented in Chap. 4 for the midpoint of a labium x:

$$\frac{dx}{dt} = y\,, \tag{5.4}$$

$$\frac{dy}{dt} = -k + (b - \beta)y - c^2 y\,, \tag{5.5}$$

which has a stationary solution defined by the conditions $dx/dt = 0$ and $dy/dt = 0$; this solution is $y = 0$ and $x = 0$. Now, what are the conditions under which this solution is *stable*? As with the stationary case, we can concentrate on the behavior of small departures ϵ_x and ϵ_y from this equilibrium, and just concentrate on a linear approximation to the complete problem. Neglecting the nonlinear terms, we obtain

$$\frac{d\boldsymbol{\epsilon}_x}{dt} = \boldsymbol{\epsilon}_y\,, \tag{5.6}$$

$$\frac{d\boldsymbol{\epsilon}_y}{dt} = -k_x + (b - \beta)\boldsymbol{\epsilon}_y\,, \tag{5.7}$$

known as the *linearized* problem. As it is a linear problem, its solutions are simple to compute. If we write it in a matrix form $d\epsilon/dt = \mathsf{DF}\epsilon$, the eigenvalues of the matrix DF allow us to gain a good insight into the behavior of perturbations around the stationary solution. For those parameter values such that the imaginary parts of the eigenvalues are different from zero, the system will behave in an oscillatory manner. However, depending on whether the real part is larger or smaller than zero, those oscillations will lead us either away from the stationary solution or towards it. For this reason, we call the curve in parameter space such that the real parts of the eigenvalues of DF are zero (while the imaginary parts may be different from zero) the bifurcation curve. In this problem, this curve is simply $b - \beta = 0$.

5.2 The Construction of Syllables

5.2.1 Cyclic Gestures

So far, we have concentrated on the mechanisms by which the labial oscillations that occur during the vocalization of a syllable are turned on and off. We are now ready to focus on the most remarkable feature of a syllable: the variation of the fundamental frequency. In terms of the mechanisms that we have described, an upsweep syllable corresponds to the oscillations being turned on at a frequency lower than that at which they are turned off. Since the frequency of the oscillations is determined by the labial restitution coefficient, a variation of the frequency will be produced by a variation of this parameter. It seems counterintuitive that the labial elasticity can be changed; it is tempting to suppose that its restitution coefficient K has a fixed value, determined by the nature of the labial tissue. However, this coefficient can be controlled. The contraction of the siringealis ventralis muscle increases

the distance between the cartilaginous rings on which the labia rest. Therefore, a contraction of this muscle increases the stiffness of the labia. This implies a higher restitution coefficient, which in turn implies an increase in the oscillation frequency.

In this way, the generation of a syllable in which the fundamental frequency varies can be achieved through the construction of a slow movement through the parameter space (of bronchial pressure and muscle tension). By this we mean that instead of there being a constant tension and pressure, the vocalization is performed while the parameters are slowly changed. "Slow" has, in this context, a precise meaning: the rate at which the parameters are changed is very small compared with the frequency of the oscillations that are turned on. Let us suppose, for example, a trajectory in the parameter space such as the one illustrated in Fig. 5.4a. As this path is traveled along, several dynamical changes will take place. As we go from point 1 to point 2, the oscillations are turned on (the interlabial pressure overcomes the dissipation in the system). As the pressure is further increased from point 2 to 3, the most important change is an increase in sound intensity and a spectral enrichment of the signal, without major changes in the fundamental frequency of the oscillations. As we go from point 3 towards point 4, as the restitution coefficient k increases and the oscillations become faster (i.e., the fundamental frequency increases). The path from point 4 to 5 is associated simply with a spectral impoverishment of the signal, which becomes progressively more harmonic and less intense, but without significant changes concerning its fundamental

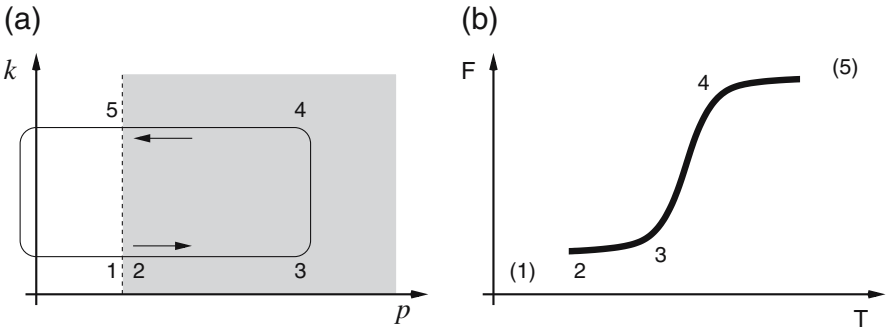

Fig. 5.4. The most remarkable feature of a syllable that has to be reproduced by a model is the changing vibratory behavior over time within a syllable, i.e., the changing fundamental frequency of the labial oscillations. For example, upsweeps and downsweeps denote syllables in which the oscillations are turned on at a lower and higher frequency, respectively, than that at which they are turned off. In order to generate an upsweep, we have to enter the region of oscillations in the parameter space with a k value smaller than the value at which we exit the region in which oscillations take places. (**a**) Path in parameter space. (**b**) Sonogram associated with the path in (**a**)

frequency. As we move out of the shaded region, just before getting to point 5, the oscillations stop [Gardner et al. 2001, Laje et al. 2002].

Now, it is ubiquitous in birdsong that certain syllables are repeated several times. This, in terms of our path, can be reproduced if, from point 5 in the parameter space, we return to point 1, and repeat the path as many times as necessary. In Fig. 5.4b, we can see a sonogram of the syllables generated (as explained in Chap. 4) using values of pressure and tension that change along the path shown. To a piece of sound analysis software, the results obtained with our equations are indistinguishable from a real recording in digital format. Therefore, we can study the result as usual. We could, for example, reproduce it in front of a bird and study the bird's behavior.

5.2.2 Paths in Parameter Space

In the previous example, the pressure started and ended at the same value after a complete cycle, and so did the labial tension. However, there was a delay between the cycles of pressure and tension. In our example, the tension started to grow after the pressure reached its maximum value. We say that the two cycles were out of phase, and we quantify this delay in terms of a quantity that we call the phase difference ϕ_0 [Gardner et al. 2001].

The simplest conceivable path in the parameter space reflecting our ideas about a "cycle" in pressure and tension and a "delay" between them can be written mathematically as

$$p(t) = p_0 + p_1 \cos(\phi(t)) \,, \tag{5.8}$$

$$k(t) = k_0 + k_1 \cos(\phi(t) + \phi_0) \,. \tag{5.9}$$

The parametrization of time through $\phi(t)$ allows us to control the speed at which the path is traversed. According to our previous discussion, $d\phi(t)/dt$ must be very small compared with the frecuency of the vocalization. However, the key parameter in (5.8) and (5.9) is the phase difference ϕ_0. Let us explore the reason.

Notice that a delay in the tension with respect to the pressure for a syllable such as the one displayed in Fig. 5.4 can be achieved with a value of ϕ_0 between ϕ and 2ϕ. How different would the syllable be if, instead of a delay in the tension with respect to the pressure, both variables evolved simultaneously? By "simultaneously" we mean *in phase*, that is, $\phi_0 = 0$. In this case, the path would look like that in Fig. 5.5. In this situation, as the pressure increases, so does the tension. Therefore, at the moment at which the oscillations start (between point 1 and point 2 in the parameter space), the frequency of oscillation increases. The process continues until both the pressure and the tension reach their maximum values. At this point the labia are oscillating with a rich spectral content and a high frequency. From this point, the pressure and tension begin to decrease (towards point 5 in the

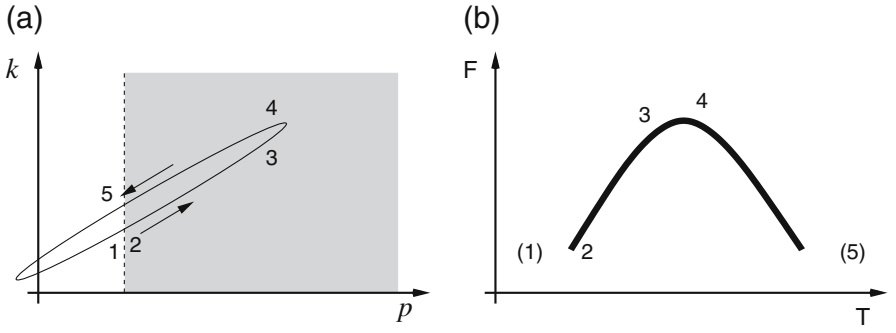

Fig. 5.5. If we enter the region of oscillations with a value of the parameter k similar to the value at which we exit the region, the syllable will have a fundamental frequency that evolves in time like a "mountain". In this case, $f_0 \sim 0$

parameter space). In this way, the syllable shows a time interval in which the frequency decreases, like one of the syllables in Fig. 5.1.

Along the same lines of analysis, we can understand how it is possible to generate a downsweep syllable. We should think of a path in the parameter space similar to the one displayed in Fig. 5.4, but traversed in reverse order. This means that the pressure should be delayed with respect to the tension, as opposed to the case shown in Fig. 5.4. This can be done by choosing ϕ_0 to have a value between zero and π. Before the oscillations start, the tension value should be high, and only after such a tension level is reached does the bronchial pressure increase to the value necessary for the oscillations to start. After this, the tension decreases. Finally, the pressure will decrease to values small enough that dissipation dominates the dynamics, and the labia will stop. The tension should increase again, if we want the cycle to repeat itself.

That the most important morphological features of a syllable can be understood in such simple terms is remarkable. In Fig. 5.6 we show several different syllables, together with the paths in the parameter space that produced them, and the corresponding phase differences. Every path was constructed according to (5.8) and (5.9). The only difference between them is the choice of the phase difference ϕ_0. Figure 5.6 says that the bronchial pressure and the labial tension vary in a cyclic way, and by simply changing the relative delay of a cycle with respect to the other we are able to change the morphology of a syllable. Of course, there are many other changes that occur simultaneously with this gesture: beak movements (which, as we saw, can modify the filtering of the signal), posture, etc. However, the basic gesture necessary to produce a syllable is a cycle in the bronchial pressure and the muscle tension.

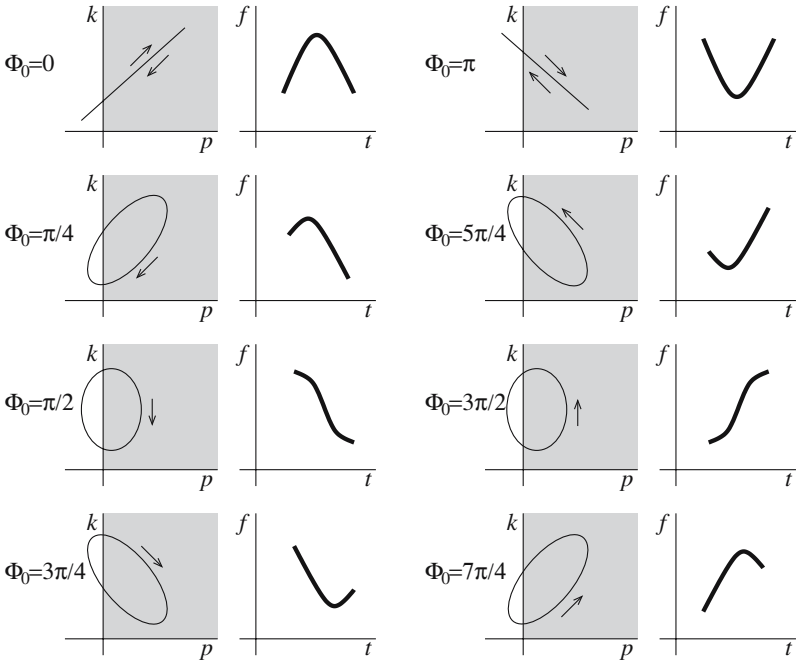

Fig. 5.6. A rich diversity of syllables can be generated by simply changing the delay between cyclic respiratory and syringeal gestures. The phase difference ϕ_0 in (5.8) and (5.9) is monotonically increased here in steps of $\pi/4$ from top to bottom, and then left to right, to generate upsweeps, downsweeps or nonmonotonic syllables

5.3 The Active Control of the Airflow: a Prediction

When we described the anatomy of the oscine syrinx in Chap. 3, we highlighted the bipartite nature of the vocal organ. However, the discussion of the labial oscillations in the previous subsections in this chapter refers to one pair of labia only. This restriction is appropriate since, some birds use mostly one side of their vocal organ. Some species use predominantly the left side (such as the waterslager canary) and others the right side (such as the zebra finch, *Taeniopygia guttata*, in order to produce calls), while some species use both sides indiscriminately (such as the brown thrasher, *Toxostoma rufus*). In some cases, both sides are used simultaneously during a vocalization. A well-studied example of this class is the last syllable in the call of the black-capped chickadee (*Parus atricapillus*), in which not only are the two sources active but also they interact acoustically [Nowicki and Capranica 1986].

In order to achieve the lateralization implied in the use of one side of the syrinx, the bird actively closes the side not used, pushing the labia against each other. In some species, the muscles in charge of performing this task are the siringealis dorsalis and the tracheobronchialis dorsalis. In the case of

the brown thrasher, these muscles rotates bronchial cartilages towards the syringeal lumen [Goller and Suthers 1996a].

The role of the muscles used to actively close the syrinx by adducting labia can be taken into account by an additional term f_0 in our simple model, (4.5). With this additional term, the model reads

$$\begin{aligned} \dot{x} &= y\,, \\ \dot{y} &= -kx + (\beta - b)y - cx^2 y - f_0\,. \end{aligned} \tag{5.10}$$

The association of the parameter f_0 with the activity of the muscles that actively adduct the labia can be seen by considering the equilibrium points of the system, also known as *fixed points*. The fixed points $\bar{x}$ of a system are found by setting all time derivatives to zero, that is, in our system,

$$\bar{x} = -\frac{f_0}{k}\,. \tag{5.11}$$

By varying the value of f_0, we can move the fixed point either towards the position representing the labia being pushed against each other or towards the position representing the labia being pushed against the walls. At these positions, oscillations are prevented.

This gesture can eventually completely adduct the labia. However, the muscle used here is briefly active on the side used to vocalize, prior to the emission of sound. Owing to this additional gesture, the pressure at which the labial oscillations effectively start is larger than that necessary if only the mechanisms described in the previous section are active. In this way, the bird delays somewhat the beginning of the oscillations. It also happens for this species that the end of the labial oscillations is anticipated by the active closing of the syringeal lumen through the dorsal muscles. It is worth pointing out that the value of the force necessary to "hold" the oscillations depends on the value of the restitution coefficient of the tissue (given by the activity of the siringealis ventralis muscle). In our simple model, the conditions under which no oscillations take place can be written as

$$|f_0| > k\sqrt{\frac{(\beta - b)}{c}}\,. \tag{5.12}$$

That is, if the labia are vibrating at a higher frequency (higher k), a larger force from the dorsal muscles is needed in order to prevent the oscillations. This is a precise prediction of the physical model described so far, and has been validated by experimental measurements [Mindlin et al. 2003].

We are now going to work with these physical models in order to reconstruct the dynamical character of the control parameters needed to reproduce artificially the syllables of a song. The result of this exercise is displayed in Fig. 5.7, in which the natural and artificial songs of a chingolo sparrow (*Zonotrichia capensis*) are displayed. In order to build the artificial song, it

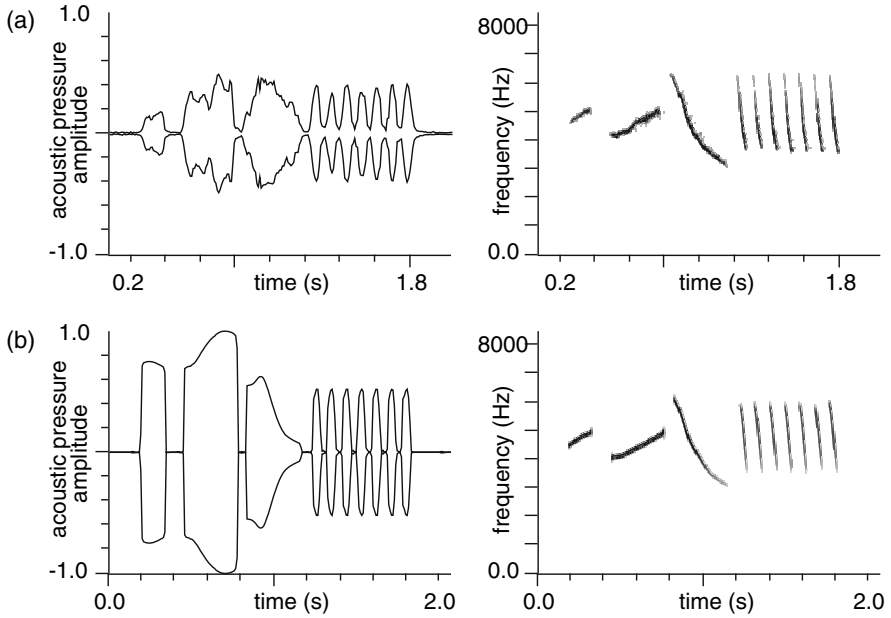

Fig. 5.7. How good are models? The simple rules described in the text allow us to reproduce the song of a chingolo sparrow (*Zonotrichia capensis*). The acoustic-pressure amplitude and the sonograms of the recorded sounds are illustrated in (**a**), while the synthetically generated sound is represented in (**b**)

was necessary to find the values of the parameters k and p, at successive instants, that would drive the model in order to generate the song. With the values displayed in Fig. 5.8, it was possible to generate a song remarkably similar to the natural one [Laje et al. 2002].

It is tempting to try one more step. Where are these gestures generated? Clearly, the activities of the muscles involved should vary in the precise way that we have described. Therefore, some sets of neurons (dedicated to conveying the instructions to the muscles) should let a coordinated firing pattern emerge. What neurons are these? Where are they located? How do they interconnect? We shall address these issues in Chap. 8.

5.4 Experimental Support

In the model discussed in the preceding section, the main acoustic parameters were correlated with the air sac pressure, the activity of the vS muscle and the activity of the dTB. In order to build confidence in this model, Franz Goller and Roderick Suthers recorded these variables. Then, these experimental records were used to drive the model. In this way, one could compare

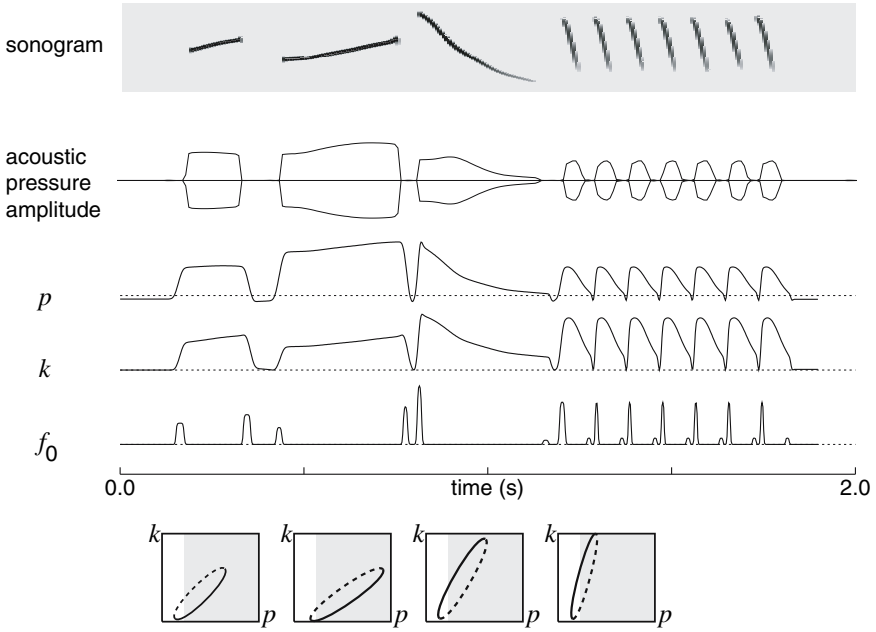

Fig. 5.8. In order to reproduce the song of the chingolo sparrow, we had to generate paths in the parameter space, which we illustrate in this figure. The sonogram of the synthetic song is shown in the first panel. The next panel shows the synthetic acoustic pressure predicted by the simple model. The panels below this one display the values of pressure, restitution constant and f_0 that had to be used in order to drive the model so it could reproduce the desired song. The last panels show the trajectories of the parameters p and k used in our simulations in a p, k parameter space. The *dashed lines* correspond to pieces of the trajectories in which the labia are not allowed to oscillate owing to the additional force f_0. The ellipses are traversed counterclockwise (modified after [Laje et al. 2002])

the recorded song with the synthetic song that was produced from the electromyographic (EMG) data that had been recorded [Mindlin et al. 2003].

The air sac pressure was recorded with a flexible cannula inserted into the anterior thoracic air sac and connected to a piezoresistive pressure transducer. The muscle activity was recorded with wire electrodes which were implanted into the syringeal muscles. These wires were made of stainless steel and were secured to the tissue with a microdrop of tissue adhesive. After the implantation of the microelectrodes, the wires were led out and routed to the back, and then the air sac was closed. Details of these procedures were published in [Goller and Suthers 1996a]. The data used in the experiments designed to test the model were simultaneous records of song, air sac pressure, dTB activity and vS activity. The data were taken from two cardinals (*Cardinalis cardinalis*).

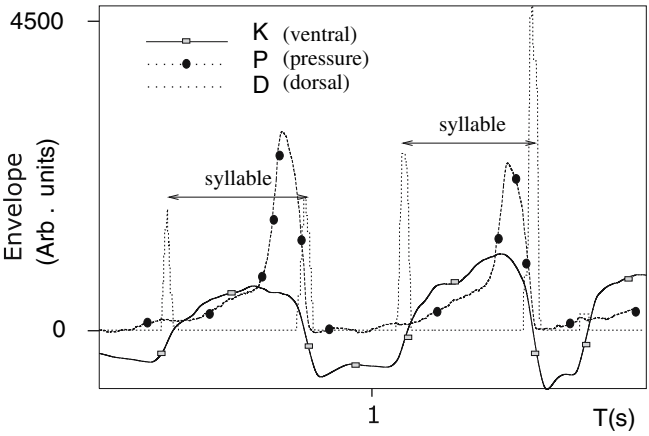

Fig. 5.9. The pressure and the smoothed–rectified activities of the dS and vS muscles during the production by cardinals of a pair of syllables. Adapted from Fig. 5.4 in [Mindlin et al. 2003]

The EMG data were sampled at 40 000 points per second. In order to reconstruct a measure of the muscle activity, the signals were rectified and smoothed. Rectification is needed in order to eliminate the artificial biphasic nature of the recordings, a result of the bipolar nature of the electrodes. The smoothing consisted of replacing each point by the average of its neighbors within a 0.2 sec window. In Fig. 5.9, we show the processed data. The vocalizations are produced basically when the air sac pressure is positive. Notice that the activity of the vS muscle increases towards the end of the vocalization. Therefore, according to the model, we expect the syllables being produced to be *upsweeps*.

The relationship between an EMG signal and the tension that is generated in the muscle is likely to be a complex one. Even from experiment to experiment, it is reasonable to expect important variations owing, for example, to the different contact areas between tissue and wire. For this reason, an absolute calibration is not possible. However, for the qualitative test that we have described, a monotonically increasing function relating the EMG to the tension is enough.

The actual test consisted of two parts. First, linear relationships were assumed between the envelopes of vS activity, dS activity and pressure (measured in Volts), and the functions k, f_0 and $\beta - b$, respectively, in (5.10). This procedure had to be repeated for each experiment. Then, these "experimental" instructions had to be used as driving functions in the model. The final step required the comparison between the syllable synthetized using the experimentally driven model, and the actual song. It is important to require in this test that within an experimental session, the same values of the fitting parameters used to map the EMG data to the driving functions (which were

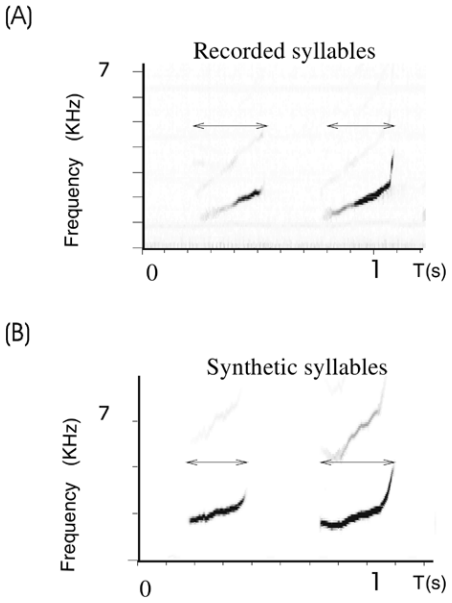

Fig. 5.10. (**A**) a sonogram of a recorded fragment of song, and (**B**) a sonogram of a synthetic song, generated when the model was driven by linear functions of the smoothed envelopes of the activities recorded during vocalizations

computed with data from one syllable) are also used for other syllables. In Fig. 5.10, we display sonograms of an actual song and one computed from EMG and pressure data.

The measurement of the activity of the dorsal muscle dS allows us to test the prediction described in the previous section regarding the relative sizes of the activities at the beginning and end of each syllable, depending on the temporal evolution of its fundamental frequency. The analysis of the model predicts that for upsweep syllables, the minimum value of f_0 needed to prevent oscillations at the begining of the syllable is smaller than the value needed to stop the oscillations at the end. The ratios between the maximum activities at the end and at the beginning of 17 syllables are displayed in Fig. 5.11. The first six syllables are upsweeps, while the rest are downsweeps. The latter show ratios smaller than one, as predicted.

The physical processes involved in birdsong are very rich, and yet it is possible to attempt to produce simple models leading to qualitative agreement with data. Since the origin of the sound in many examples of birdsong has been traced to the vibration of labia, the existence of simple oscillatory models reproducing the sounds is only natural. However, it is not trivial even in principle to interpret model parameters in terms of the actual biological

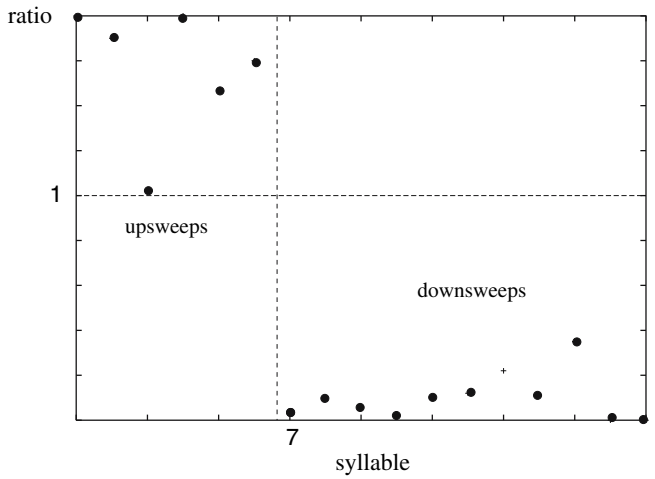

Fig. 5.11. The ratios between the local maximum activity of the dS muscle at the end and at the beginning of a vocalization. Seventeen syllables from two birds were analyzed, the first six being upsweeps. Adapted from Fig. 5.8 in [Mindlin et al. 2003]

forces. The experiments described in this section, however, build confidence in the physical model that we have presented in this chapter.

Much of the research on birdsong focuses on the neural processes involved in the learning of the vocalizations. By illuminating the mechanisms by which the activities of the syringeal muscles control the syrinx, we can help to build a bridge between neural activity and vocal behavior. An implication of this modeling effort for neurophysiologists in the field is that in order to map central neural activity onto peripheral motor activity, it is important to relate neural activity to the driving parameters rather than to the acoustic properties of the emitted song [Suthers and Margoliash 2002].

5.5 Lateralization

The functional lateralization of the brain remains an open problem. In birdsong production, this phenomenon is present [Allan and Suthers 1994]. The degree of asymmetry present in the behavior ranges from low to very high. In some species, there is a clear unilateral dominance (only one side of the syrinx is used), while in others, there are equal contributions from both sides [Suthers 1990]. For a long time, it was very puzzling that this functional asymmetry did not have a clear anatomical correlation in the song control nuclei. This problem was addressed by Goller and Suthers [Goller and Suthers 1995], who studied the activity in the muscles controlling both sides of a brown thrasher's syrinx. As we have seen, some muscles are involved in the control of the labial tension (and therefore of the frequency of the vocalization),

and others are involved in the gating of the airflow. The findings reported in [Goller and Suthers 1995] indicate that it is only in the case of the gating muscles that the activity is lateralized. The other syringeal muscles are active on both sides of the syrinx.

According to our models, the vocalization is largely determined by the phase differences between the gestures affecting tension and pressure. It is then natural to ask about lateralization of the muscle activity involved in the control of the air sac pressure. In [Goller and Suthers 1999], it was investigated whether the activity of respiratory muscles reflected the lateralized activity of the vocal organ during the production of birdsong. Those authors found that the activity was always present, independently of whether the song was produced by one or both sides of the syrinx.

6 Complex Oscillations

To describe the songs of birds using the songs of the canary and the chingolo as the only examples might seem suspicious. We are all familiar with the wide range of timbre that birds can show in their songs. Can they all be described in terms of the simple mechanisms presented so far? The answer is negative, as we shall discuss in this chapter. Unavoidably, we shall have to introduce more subtle effects in order to widen the spectrum of phenomena that are presented by different species and under different circumstances.

6.1 Complex Sounds

6.1.1 Instructions vs. Mechanics

The spectral description of the song of some birds reveals the existence of phenomena more complex than the ones studied so far. In Fig. 6.1, two sonograms are displayed. They correspond to the songs of the eastern slaty thrush (*Turdus subalaris*) [Straneck 1990a] and the grassland sparrow (*Ammodramus humeralis*) [Straneck 1990b]. They are more complicated than the sonograms described in the previous chapters.

In Fig. 6.1a, above the curve representing the fundamental frequency, there is a second curve which has little in common with the first curve. It neither begins nor ends when the first curve does, and the frequency values do not correspond to multiples of those of the first curve, i.e., the second curve does not represent a harmonic of the first one. The completely different way in which these curves behave is a signature of the fact that the bird has used both sources in an independent way in order to generate its vocalization.

Compared with Fig. 6.1a, Fig. 6.1b is misleadingly simple. Notice that there is a qualitative change in the shape of the sonogram, indicated by the arrow. The line that corresponds to the time evolution of what was the lowest frequency at the beginning of the vocalization appears to be accompanied after some time by new curves at its sides. Have we seen sonograms like this one before? No, since these new "parallel" curves are not harmonics of the previous fundamental frequency. Recall that the harmonic frequencies are lines parallel to the curve representing the fundamental frequency (the lowest curve) that are located at multiples of the fundamental frequency. In

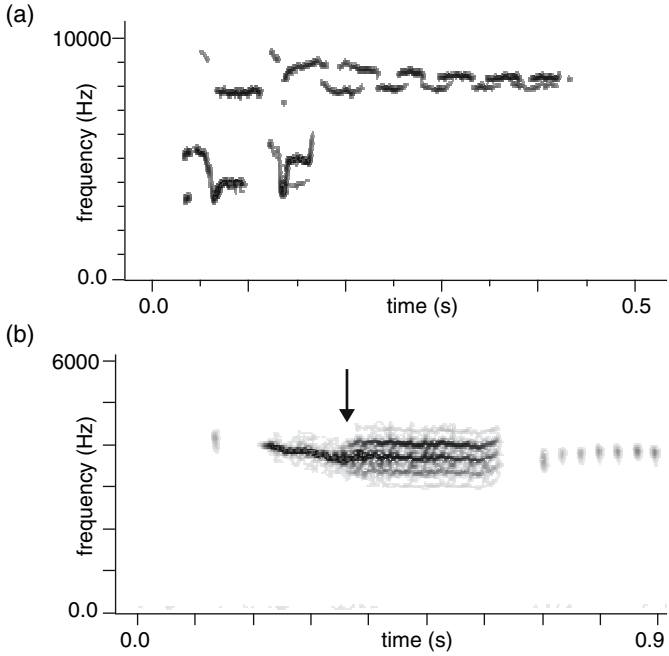

Fig. 6.1. In (**a**), we display the sonogram of a song produced by the eastern slaty thrush (*Turdus subalaris*), which uses both sides of the syrinx to produce a rather complex sound. In part (**b**), we find in the sonogram of the grassland sparrow's song (*Ammodramus humeralis*) a series of paralell lines, which are a signature of rather complex dynamics of the labia used to produce the sound. This sonogram corresponds to a sound which cannot be reproduced with the simple models described so far

this case, the lowest frequency before the transition is approximately 4000 Hz; the first harmonic should therefore be at approximately 8000 Hz. However, the new lines appear at 3800 Hz and 4200 Hz, etc.

Both examples imply a higher degree of complexity of the sound signal than in the ones displayed in the previous chapters. However, there is an important difference between them. In the first case, the bird uses both sides of its bipartite vocal organ. The sound is generated by two sources that each contribute their fluctuations (in principle, in an independent way), and each contribution can be understood in terms of the elements that we have presented so far. It is certainly true that to activate both sound sources with a certain degree of coordination (instead of keeping one silent through the active closing of the lumen) implies an important degree of complexity at the level of neural the instructions sent to the vocal organ. However, the physics involved is as simple as it was before.

The second case, in contrast, is more subtle: although the dynamics of the lines appear to indicate that there is just one active source of sound,

the spectrum reveals the existence of a physical phenomenon that we cannot interpret with the elements presented so far. In summary, we have two samples of complex sounds, for which the origins of the complexity are quite different. It is natural, then, to wonder if the complexity of a vocalized sound is necessarily a signature of the complexity of the neural instructions sent to the vocal organ by the brain. The first researchers who became interested in this problem described the possible physical nontrivial behaviors of syringeal labia subjected to a simple constant airflow [Fee et al. 1998]. The study was carried out in vitro on syringes surgically extracted from zebra finches (*Taeniopygia guttata*). The degree of complexity of the sound signal produced in these experiments was the same as that found in the natural vocalizations of the bird. In this way, it was established that there is a degree of complexity associated with the mechanics of the vocal organ, beyond the richness of its neural control.

6.1.2 Subharmonics

We presented the sonogram in Fig. 6.1b as an example of a complex sound signal. Why? Remember (Chap. 1) that a signal of a given fundamental frequency F repeats itself every T time units (T is the period associated with the fundamental frequency; $T = 1/F$). It is possible for a signal to undergo a modification of its spectral content and acquire or lose harmonics (for example, after being filtered by the vocal tract), but it will always repeat itself after T time units. The harmonics, which are multiples of the fundamental frequency, make an important contribution to the timbre of the sound, but they do not modify its periodicity, since they repeat themselves every T/n units of time, where $n = 1, 2, 3, \ldots$. In contrast, the appearance of *subharmonics* (that is, frequencies that are submultiples of the fundamental frequency) corresponds to signals that repeat themselves not after T but after nT time units. In the language of Chap. 1, a time function such as

$$y_1 = A\cos(\omega t) + \epsilon_{\text{supra}}\cos(2\omega t) \tag{6.1}$$

has a period $T = 2\pi/\omega$, no matter the value of ϵ_{supra}. Its sonogram will consist of a fundamental frequency $F_1 = 1/T$ with amplitude A plus its first harmonic $F_2 = 2F_1 = 2/T$ with amplitude ϵ_{supra}. In contrast, a function such as

$$y_1 = A\cos(\omega t) + \epsilon_{\text{sub}}\cos\left(\frac{1}{2}\omega t\right) \tag{6.2}$$

has a period $T_{\text{new}} = 2T$ (twice as large) as soon as ϵ_{sub} is different from zero, since $2\pi/T_{\text{new}} = \omega/2 = (1/2)(2\pi/T)$. What mechanisms can give rise to such a behavior?

6.2 Acoustic Feedback

6.2.1 Source–Filter Separation

One of the possible ways of obtaining oscillations of this sort, without abandoning the model that we introduced in Chap. 4, is through *acoustic feedback*. When we analyzed the onset of oscillations for the syringeal labia, we mentioned that when the labia have a divergent profile, the average interlabial pressure is low with respect to the bronchial pressure. The reason is that when the profile is divergent, the average pressure between the labia is similar to the tracheal pressure (and similar in turn to atmospheric pressure). We also discussed the existence of pressure perturbations at the input of the vocal tract as a result of adding the flow perturbations injected by the sources to the perturbations induced by reflected pressure waves. In that discussion, we assumed implicitly that the existence of pressure perturbations in the vocal tract did not affect the labial dynamics. The theory that assumes that the labial dynamics are independent of whatever happens in the filter is known as the *source–filter theory*. In many cases, as during normal human speech or the song of most bird species, this hypothesis is a most sensible one. But source–filter separation does not hold in every case, or at least we cannot a priori affirm that it holds.

6.2.2 A Time-Delayed System

Formally, the dynamics of a labium are described in terms of its midpoint position x by

$$M\ddot{x} + B\dot{x} + Kx = P_g\,, \tag{6.3}$$

where P_g is the average interlabial pressure, given by

$$P_g = P_i + (P_s - P_i) f(x, \dot{x})\,; \tag{6.4}$$

$f(x, \dot{x})$ is a generic function, which, in the flapping model, is $f(x, \dot{x}) = 1 - a_2/a_1$ (see Chap. 4). The source–filter approximation assumes that the pressure at the entrance of the tract P_i is zero (that is, atmospheric pressure). A more realistic approximation takes into account the fact that this pressure is the result of adding perturbations due to the modulation of the airflow and the perturbations arriving back after being reflected, mostly at the beak:

$$P_i(t) = s(t) + P_{\text{back}}(t - \tau)\,, \tag{6.5}$$

$$P_{\text{back}}(t) = -\gamma P_i(t - \tau)\,, \tag{6.6}$$

where τ is the time it takes for a sound wave to travel the length of the tract, and γ is the reflection coefficient at the beak. This equation expresses the fact that the pressure perturbations at the input of the tract are the result of adding the perturbations induced by the labial fluctuations ($s(t)$)

and the pressure wave arriving as a result of reflection at the end of the tract. If these perturbations are $P_{\text{back}}(t)$ at time t, the total perturbation is the result of adding $s(t)$ and $P_{\text{back}}(t-\tau)$, since it takes the pressure wave traveling backwards a time τ to reach the opposite end of the tract. The source of pressure perturbations $s(t)$ is a function of all the variables describing the dynamics of the labium. As discussed in Chaps. 1 and 2, the pressure perturbations $s(t)$ at the entrance of the vocal tract are generated by time variations of the flow $U(t)$ injected by the glottis, which in turn is a function of the variables x and $\dot{x}$.

The introduction of a time delay into the differential equation describing the labial dynamics is responsible for a dramatic change in the kind of solutions that can be expected. In order to gain an intuitive idea about this observation, we can compare the simple dynamics obtained from the system

$$\frac{dN(t)}{dt} = -\frac{\pi}{2\tau}N(t) \tag{6.7}$$

with the solution of

$$\frac{dN(t)}{dt} = -\frac{\pi}{2\tau}N(t-\tau)\,. \tag{6.8}$$

In the first case, we obtain a linear decay to zero, $N(t) = A\exp(-\pi t/2\tau)$, while in the second case the system evolves harmonically in time: $N(t) = A\cos(\pi t/2\tau)$.

There are two parameters, then, that are important in order to obtain complex dynamics within this framework: one (the *coupling parameter*) measuring the amplitude of the perturbation injected by the glottis at the entrance of the vocal tract, and one describing the delay involved (which can be taken simply as the time it takes a sound wave to travel the length L of the vocal tract). For large enough values of the coupling parameter, the pressure fluctuations that are established at the base of the trachea can be important and affect the labial dynamics [Laje et al. 2001]. The vocal-tract input pressure values found by the labia in two consecutive situations in which the profile is divergent (i.e. opened to the trachea) are not going to be equal. Whenever this is the case, the labial oscillations can be more complex than expected, and the period of the signal (if it exists) can be very large. This is a possible mechanism for the appearance of low frequencies (subharmonics) in the spectral analysis of a signal.

6.2.3 Coupling Between Source and Vocal Tract

An expression for the coupling parameter is needed in order to continue the analysis. In other words, we need to develop an expression for $s(t)$ in (6.5) in terms of the variables x and $\dot{x}$. The coupling parameter (that is, the coefficient hopefully appearing in front of the variables in the expression for $s(t)$) will in general be a combination of anatomical parameters of the syrinx and vocal

tract (such as lengths, widths and cross sections) and parameters related to sound (such as the frequency and speed of the sound, and the air density).

A rigid-walled waveguide can propagate only plane waves if its cross-sectional dimensions are much smaller than the sound wavelength [Kinsler et al. 1982]. This is the case for the avian trachea. In the species we are considering, and as a first approximation, the trachea is a very narrow tube around 2 cm long and 1 mm wide, with a *cutoff frequency* (that is, the frequency below which every propagating wave is a plane wave) of around 100 kHz. We can then assume that the sound waves in the avian trachea are plane waves, and therefore make use of the very simple boundary conditions (6.5) and (6.6) to account for the superposition of forward-emitted and backward-reflected waves.

However, we do not expect the sound wave to be a plane wave very near the source. As it is emitted by the glottis, the sound wave is of a rather diverging nature; for the sake of simplicity, we assume that the glottis behaves as a local emitter of *spherical* diverging waves. Recall now our discussion of the concept of flow in Chap. 2. Time variations of the flow $U(t)$ (which is related to the air particle velocity v through $U = vA$) induce pressure perturbations $s(t)$ at the vocal-tract input. For a spherical sound wave, assuming that the pressure perturbations $s(t)$ and air particle velocity $v(t)$ are both harmonic functions of time, the relationship between them can be written as

$$s(t) = zv(t)\,, \tag{6.9}$$

where $z = R + i\omega I$ is the *complex specific acoustic impedance*. R and I are called the *specific acoustic resistance* and the *inertance*, respectively:

$$R = \rho c\,\frac{(kd)^2}{1+(kd)^2}\,, \tag{6.10}$$

$$I = \rho\,\frac{d}{1+(kd)^2}\,, \tag{6.11}$$

where c and ρ are the sound speed and the air density, and d is the anatomical parameter shown in Fig. 6.2. Note that both R and I depend on the frequency f of the sound wave through the wavenumber $k = 2\pi f/c$. Note also that R and I are not independent quantities; they are related through $R = 4\pi^2 f^2 (d/c)\, I$. A plot of R and I as a function of the frequency f is shown in Fig. 6.3.

The parameters R and I are the coupling coefficients we are looking for (v can be written in terms of the variables x and $\dot{x}$ describing labial motion). The key issue is that the pressure perturbations $s(t)$ have two different contributions: the first one in phase with the flow, scaled by R, and the second one in phase with flow derivative and scaled by I (recall the discussion leading to (2.12)). The nature of the coupling depends on the balance between these two contributions ("resistive" coupling when the term containing R dominates, and "inertive" coupling otherwise), which in turn depends on the frequency

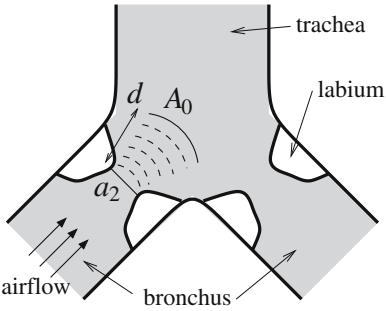

Fig. 6.2. Schematic section of the syrinx and vocal-tract input. Airflow coming from the lungs is modulated by labial oscillation. As a result, a diverging sound wave is injected at the input of the vocal tract. Dashed lines indicate the region of spherical propagation. After this region and once in the trachea (which is a very narrow tube around 1 mm wide), the sound wave is propagated as a plane wave. The parameter d is a measure of the distance between the labia and the region of plane-wave superposition

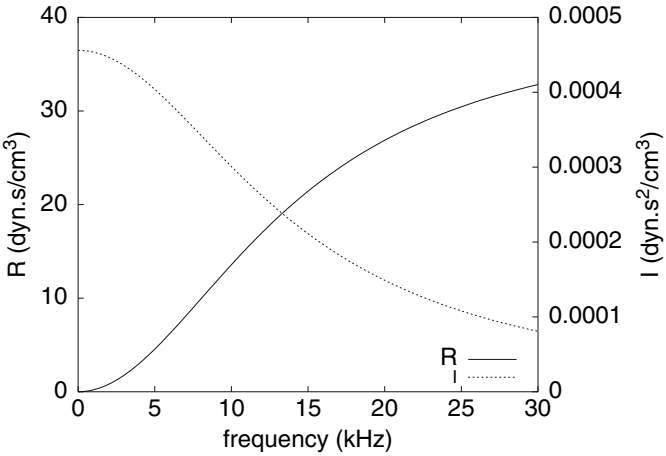

Fig. 6.3. Coupling between source and vocal tract. The dependence of $s(t)$ on the flow and the flow derivative is mediated by the specific acoustic resistance and the inertance R and I, respectively. The nature of the coupling (more in phase with either the flow or its time derivative) is governed by the fundamental frequency of the vocalization

f of the vocalization. Note in Fig. 6.3 that the coupling is inertive at low frequencies, but that it becomes resistive at high frequencies; the frequency at which the coupling changes behavior depends on the anatomical parameter d and the sound velocity c.

An important assumption in the previous discussion is that the two valves modulating the airflow are close to the junction of the trachea with the

bronchi. If this is not the case, the portion of the bronchus between the trachea and the valve will be seen by the valve as a pipe. Fletcher and Tarnopolsky computed the bronchial resonance associated with this anatomy [Fletcher and Tarnopolsky 1999]. These formants have been identified by Suthers [Suthers 1994] in a study of the oilbird (*Steatornis caripensis*). He identified frequencies which could be associated with the bronchial segments, and showed that they were behaving as quarter-wave pipe resonators.

6.3 Labia with Structure

6.3.1 The Role of the Dynamics

There are other mechanisms that allow us to account for the spectral complexity described above. But we shall have to enrich our model of the labial oscillations in order to address them. In our discussion in Chap. 4, the upper part of a labium moves in a way that is perfectly correlated with the lower part. This is due to our description of the labial movement in terms of a superposition of two simple modes. The first mode consisted of a global displacement, and the second one of an upward wave, together making up a kind of flapping motion. In this situation, the upper and lower parts of a labium move a little out of phase, the upper part always following the lower part.

A model that allows for richer behavior can be built if the upper and lower parts of each labium are thought of as two masses attached to each other by means of a spring. In such a model, the labia can show more complex behavior, beyond the behavior described above [Herzel et al. 1995, Herzel et al. 1996]. This was the strategy followed by Fee et al. [Fee et al. 1998] to explain the existence of subharmonics in the song of zebra finches (*Taeniopygia guttata*). The interesting aspect of this observation is that the acoustic properties of the song are not uniquely controlled by the nervous system of the bird. The mechanical properties of the vocal organ can contribute to generating new properties of the sound when a small change in the control parameters is made. In fact, the work by Fee et al. was performed on a syrinx in vitro. Therefore, the appearance of subharmonics for certain values of the pressure does not require delicate muscular control. The model used by Fee and coworkers, known as the "two-mass model" (first proposed in [Ishizaka et al. 1972]) and illustrated in Fig. 6.4, is capable of generating fluctuations in the airflow of a higher complexity than those generated by the simple model (4.5). The reason is that the motion of the upper part of the labium, represented by the mass m_2 in the figure, is affected (due to elastic coupling) to the lower part. This allows solutions different from the ones analyzed so far, in which a small distance between the upper parts of the opposed labia implied a large separation between the lower parts.

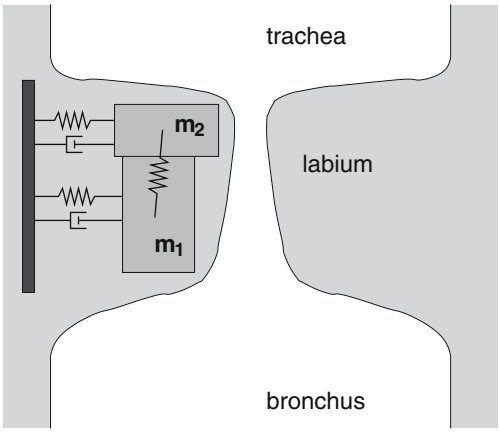

Fig. 6.4. The two-mass model of the labia. The hypothesis in the simple model described in Chap. 4 is that only two modes of oscillations are needed to account for the dynamics of the labia. A more subtle model would require us to allow different parts of the labia to display their dynamics without being constrained to generate an upward wave. The solutions of the models described so far can be generated by this new model of labial oscillation, but now, new and richer solutions can also be found

6.3.2 The Two-Mass Model

The set of hypotheses for the values of the interlabial pressure in this new model is compatible with the set corresponding to the flapping model. In the two-mass model, it is assumed that, if the labia form a convergent profile (the lower parts of the opposing labia are more separated than the upper parts), the pressure between the upper parts of the opposing labia is basically the vocal-tract pressure. On the other hand, the pressure between the lower parts of the opposing labia is a fraction of the bronchial pressure. This occurs because in this situation, the airflow is almost laminar (by laminar, we mean that the air particles travel in an orderly fashion, following smooth trajectories over the entire space available). In this regime, the pressure of the air decreases as the fluid velocity increases. Moreover, the air will travel faster as it moves from the bronchi to the interlabial space, since the passage is narrower between the labia. If there is no accumulation of air, this should further increase the air velocity. This is why the pressure between the lower parts of the labia will be a fraction of the bronchial pressure.

Let us now pay attention to the hypothesis for the case in which the labia form a divergent profile (i.e., the labia are "opened" to the trachea). It is considered that the air acquires a velocity in its passage through the space between the lower parts of the labia such that a laminar regime cannot be established in the region between the upper parts. On the contrary, a jet is established, so that the actual value of the distance between the upper parts

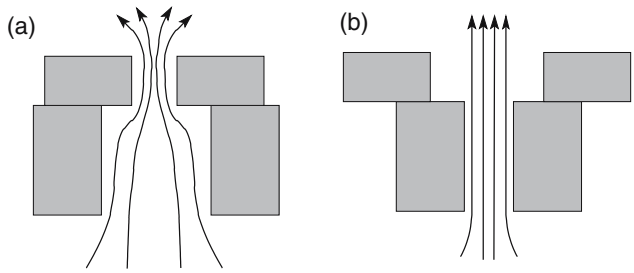

Fig. 6.5. Two snapshots of the labial motion in the two-mass model. In this model it is assumed that, if the labia form a convergent profile (the lower parts of the opposing labia are more separated than the upper parts) as shown in (**a**), the pressure between the upper parts of the opposing labia is basically the vocal-tract pressure. On the other hand, the pressure between the lower parts of the opposing labia is a fraction of the bronchial pressure. If the labia form a divergent profile as in (**b**), it is considered that the air acquires a velocity in its passage through the space between the lower parts of the labia such that a laminar regime cannot be established in the region between the upper parts. On the contrary, a jet is established, so that the actual value of the distance between the upper parts of the labia is irrelevant

of the labia is irrelevant. The interface of interest is now the one between the trachea and the lower part of the labia. For reasons of continuity, the interlabial pressure can be assumed to be equal to the tracheal pressure. These hypotheses are illustrated in Fig. 6.5.

The global result of this set of hypotheses is not much different from that in the case of the flapping model: in both models, the computation of the average pressure between the labia for a divergent configuration gives a smaller value than that for a convergent one, thus allowing a net energy transfer from the airflow to the labial oscillation. The advantage is that if we use a model in which the upper and lower parts of the labia are elastically connected instead of having perfectly correlated positions, richer motions are possible.

The results obtained by studying this model are surprisingly good if we consider the number of approximations that have been made. Why can we say that the results are reasonable? First, the equations are stated in such a way that, given the values of the pressure, the masses that approximate the labia, the restitution constants of the assumed coupling springs, etc., we are able to compute the motion of the masses. The equations are simply Newton's laws applied to this problem:

$$M_1\ddot{x}_1 + B\dot{x}_1 + K_1x_1 + K_r(x_1 - x_2) - G_1 = F_1\,, \tag{6.12}$$

$$M_2\ddot{x}_2 + B\dot{x}_2 + K_2x_2 + K_r(x_2 - x_1) - G_2 = F_2\,, \tag{6.13}$$

where $x_{1,2}$ are the departures from equilibrium of the lower and upper masses $M_{1,2}$, respectively; $K_{1,2}$ are the restitution constants for each mass; K_r

denotes the elastic coupling constant between the two masses; and $B_{1,2}$ are the linear dissipation constants. Unlike the case for the flapping model, a collision between labia is modeled here by $G_{1,2}$, an additional restoring force:

$$G_i = \begin{cases} g_i a_i \text{ during collision}\,, \\ 0 \quad \text{otherwise}\,. \end{cases} \tag{6.14}$$

Here $a_i = 2L_g(x_0 + x_i)$ denotes the lower ($i = 1$) and upper ($i = 2$) labial areas, which take negative values during a collision. L_g is the length of the glottis, and x_0 is the labial rest position. Finally, the forces $F_{1,2}$ are defined by $F_{1,2} = L_g d_{1,2} P_{1,2}$, where $d_{1,2}$ are the thicknesses of parts 1 and 2, and $P_{1,2}$ are the pressures acting on masses 1 and 2. Using Bernoulli's law, when the glottis is open we can write

$$P_1 = P_s - \frac{\rho}{2}\left(\frac{U}{a_1}\right)^2\,, \tag{6.15}$$

$$P_2 = P_0\,, \tag{6.16}$$

where $U = a_{\min}\sqrt{2(P_s - P_0)/\rho}$ is the velocity of air through the glottis ($a_{\min}$ is the minimal labial area between a_1 and a_2; it is set to 0 during a collision). These equations reflect the fact that the pressure between the upper labia is always the pressure at the entrance of the tract P_0, and the pressure between the lower labia can be either P_0 when the profile is divergent or a larger quantity (between P_0 and P_s) when the profile is convergent.

Once these rules have been stated, the important questions are: Will the masses begin to oscillate for realistic values of the pressure? Are the frequencies of the oscillations comparable to the expected ones? When this model was studied by Ishizaka and Flannagan [Ishizaka et al. 1972] in the framework of an analysis of the human voice, surprisingly good results were obtained. The model was afterwards studied extensively by many authors. (In any case, no one considers a theory to be more than a set of partial truths, ready to be abandoned for another one that explains both whatever could be explained by the previous theory plus new phenomena which could not... In fact, there is nothing as frustrating as a theory that is too successful...!) A detailed analysis of this model performed by Steinecke and coworkers [Steinecke and Herzel 1995] has shown that for sublabial pressures larger than a certain critical value, oscillations are established which are very similar to the ones predicted by our first model, that of the flapping motion. Similar in what sense? That the motions of the upper and lower parts are highly correlated, the upper part always oscillating with a certain phase difference with respect to the lower part.

6.3.3 Asymmetries

The old model in Chap. 4 could not account for the existence of subharmonics. In terms of sonograms, what the old model cannot reproduce is the lines

close to and parallel to the line of the fundamental frequency in Fig. 6.1b. The interesting thing about this complex model, involving so many masses and springs to emulate the labial motion, is that it gives us room to study what happens when we change somewhat the symmetry of the problem. Now we can afford to play with the possibility that the right and left labia might not be exactly equal. This allows us a greater degree of realism, since the bipartite nature of the syrinx does not allow, for each half, the same degree of symmetry between labia as what is present in the case of human vocal folds. Steinecke et al. [Steinecke and Herzel 1995] studied what kinds of oscillations exist for different values of the pressure and for various degrees of asymmetry in this model.

There are wide ranges of pressure for which (despite the asymmetries) the solutions are regular, but for sublabial pressures larger than a certain value – which depends on the degree of asymmetry of the problem – the solutions become very complex. In Fig. 6.6, we compare the syringeal flows that are predicted by the model for two different values of the pressure (and a fixed degree of asymmetry). In the case illustrated, the values of the masses and the coupling constant between them on the left side are 52% of the values for the right side. As can be seen in the figure, for the first pressure condition, the airflow is regular, with a certain period. For the second condition, the airflow does not repeat itself until a time lapse three times larger than in the previous case, despite the fact that the change in sublabial pressure was very slight. The computed spectra denote the additional complexity of the airflow. The origin of this complexity is in the dynamics of the masses, which can in fact become completely aperiodic for appropriate values of the sublabial pressure

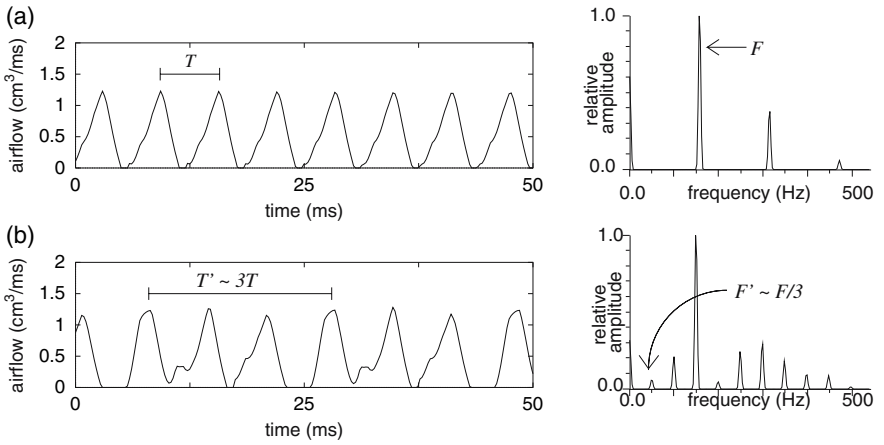

Fig. 6.6. Assymetric two-mass model showing dramatic changes in flow periodicity as the pressure parameter is changed slightly. The masses and coupling constants for the left side are 52% those for the right-side. (**a**) "Normal" oscillation with frequency F. (**b**) When the pressure is raised 10% above the value in (**a**), the frequency of the oscillation suddenly changes to $F/3$

and degree of asymmetry. And how do these pressure waves sound when the labial motion is complex? Well... not very nice. It is interesting that the presence of subharmonics can be found in juvenile birds and also in the crying of human babies. It is tempting to speculate on the evolutionary advantages of irritating mom and dad to get their attention [Robb 1988]. Can physics be an ally in this process?

6.4 Choosing Between Two Models

In Fig. 6.1b, we were presented with a most interesting and puzzling effect: a complex sound whose sonogram displayed evidence of subharmonicity. However, we have discussed two different mechanisms which are capable of accounting for this effect: coupling between the sources and the tract (Sect. 6.2), and a complex spatial excitation of the oscillating tissues (Sect. 6.3). Is there a way to find the mechanisms at play for a particular bird? The unveiling of the physical processes involved in each of these different mechanisms allows us to predict the outcome of experiments designed to allow us to choose between these competing models. In particular, we shall investigate what could be expected to occur if a bird producing complex vocalizations sings in a heliox atmosphere (i.e., one in which nitrogen, comprising 80% of ordinary air, is replaced by the less dense helium). Experiments of this sort has been used to show that some birds actively coordinate the passive-filter characteristics of their vocal tract with the output of the syrinx [Nowicki 1987] and to discern between competing proposed mechanisms for song production [Ballintijn and ten Cate 1998].

According to (6.9), the pressure fluctuations induced by the source at the input of the tract depend on the air particle velocity v and its derivative dv/dt through the coefficients R and I. The air velocity at the input of the tract can be written as $v(t) = v_0 a_2(t)/A_0$ (see Fig. 6.2) owing to mass conservation, where $v_0 = \sqrt{2P_s/\rho}$. This defines a dependence of the air particle velocity on the density of the medium. On the other hand, the coefficients R and I depend on the density, as shown in (6.10) and (6.11). Therefore, the pressure perturbation $s(t)$ can finally be written as

$$s(t) = \alpha f(x, dx/dt) + \beta \, df(x, dx/dt)/dt \,, \tag{6.17}$$

where $f(x, dx/dt)$ is a function that depends on the kinematics of the labia, and α and β scale with the density as

$$\alpha \sim \rho^{1/2} c^{-1} \sim \rho \,, \tag{6.18}$$

$$\beta \sim \rho^{1/2} \,. \tag{6.19}$$

Since a heliox atmosphere has a density 33% that of ordinary air, we obtain

$$\alpha_{Heliox} = 0.3 \alpha_{air} \,, \tag{6.20}$$

$$\beta_{Heliox} = 0.5 \beta_{air} \,, \tag{6.21}$$

which allows us to predict qualitative changes in the spectral features of syllables produced in these atmospheres. Performing numerical simulations of the models defined in the previous sections, we have obtained the syllables displayed in Fig. 6.7.

In Fig. 6.7 we display syllables obtained with models that take into account the coupling between the source and the filter (i.e. the acoustic-feedback model of Sect. 6.2) (Fig. 6.7a), and the spatial structure of the oscillating labia (i.e., the two-mass model of Sect. 6.3) (Fig. 6.7b). In each case, we compare the syllables that are generated in two atmospheres of different density (air and heliox). In Fig. 6.7b, the syllable is generated while the bronchial pressure is raised, in order to display the appearence of subharmonic behavior. The main conclusion that we can obtain from this numerical experiment is that the subharmonic behavior due to source–filter coupling can be suppressed if the density of the atmosphere is lowered. On the other hand, the changes

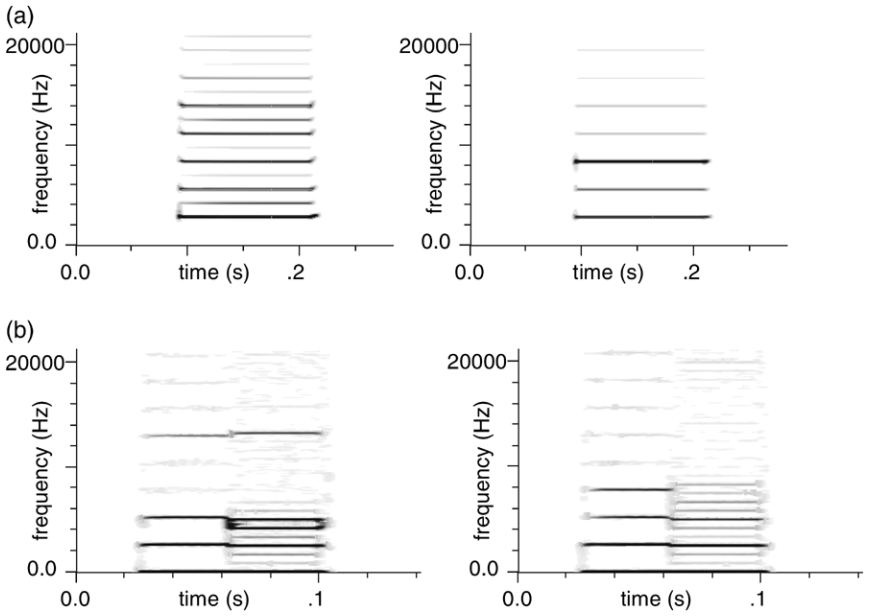

Fig. 6.7. Predicted sonograms in heliox experiments. (**a**) Acoustic-feedback model (Sect. 6.2). *Left*: a subharmonic sonogram in ordinary air. *Right*: when ordinary air is replaced by heliox, which has a density one-third that of air, the coupling between the source and tract decreases according to (6.20) and (6.21), and the subharmonic behavior disappears. (**b**) Two-mass model (Sect. 6.3). The results of the two-mass model do not depend on the air density, as long as the coupling between the source and tract is negligible. A syllable is displayed which shows a transition to a period-three solution as the sublabial pressure is slightly raised. This transition is due to the appereance of higher-order modes of labial vibration. The only difference between ordinary air (*left*) and heliox (*right*) is a change in the vocal-tract resonances

in the spectral features of the vocalization produced by structured labia are more subtle: the complexity of the oscillations at the source is not decreased, but changes are observed in the resonances, which grow by up to 75% in the heliox atmosphere.

6.4.1 Signatures of Interaction Between Sources

There is yet another surprise behind the bipartite structure of the syrinx. In the sonogram in Fig. 6.1a, we saw the signature of two sources that were simultaneously active. The sonogram consisted of two lines with independent time evolutions. There are no major difficulties in understanding this phenomenon: the sources act independently, and both inject pressure fluctuations that are added at the tracheal entrance [Nowicki 1997]. The tract will simply filter the harmonic components as we have analyzed so far, in order to achieve the final vocalization. However, when we analyzed the behavior of one source, we noticed that in some circumstances, the resulting pressure at the input of the trachea could show fluctuations so large that the labial dynamics were affected. It is natural, then, to expect that it could be possible that the fluctuations induced by an active source could influence the behavior of the second source, if this were also to become active. Since it is not likely that both sources will show oscillations at exactly the same frequency, the resulting behavior in the case of coupling can be easily recognized.

To give us an idea of what we can expect to find in the case of two coupled sources, let us revisit the simple model of the oscillating labia. Remember that the key for establishing oscillations in one source is the interlabial pressure, which is responsible for the force that overcomes the system's dissipation and sustains the oscillations. However, the existence of a coupling changes this scenario somewhat. Imagine that one of the sources is "on", and generates important pressure fluctuations at the input of the trachea. The second source, as it begins to oscillate (for example, as the activity of the dorsal muscle keeping the labia together is decreased), finds a complex scenario. Each time that the labia have a divergent profile (i.e. open to the trachea), the source will experience a fluctuating pressure that oscillates at the frequency of the first source. Therefore, the second source will oscillate with an amplitude modulated at the first source's frequency. In Fig. 6.8a, we show a harmonic signal and its spectrum (composed of a unique frequency). In Fig. 6.8b, we show the resulting signal if we modulate the amplitude of the first source with a different frequency. The spectrum of this signal now shows new frequencies, with the signature of the coupling being frequencies equal to the sum and difference of the original frequencies. Coupling between two avian sound sources has been experimentally demonstrated in a characteristic syllable of a black-capped chickadee (*Parus atricapillus*), a sort of internal duet between the two sound sources [Nowicki and Capranica 1986].

An amazing fact is that some species have the capacity to choose between using the two sources independently or making them interact. This seems to

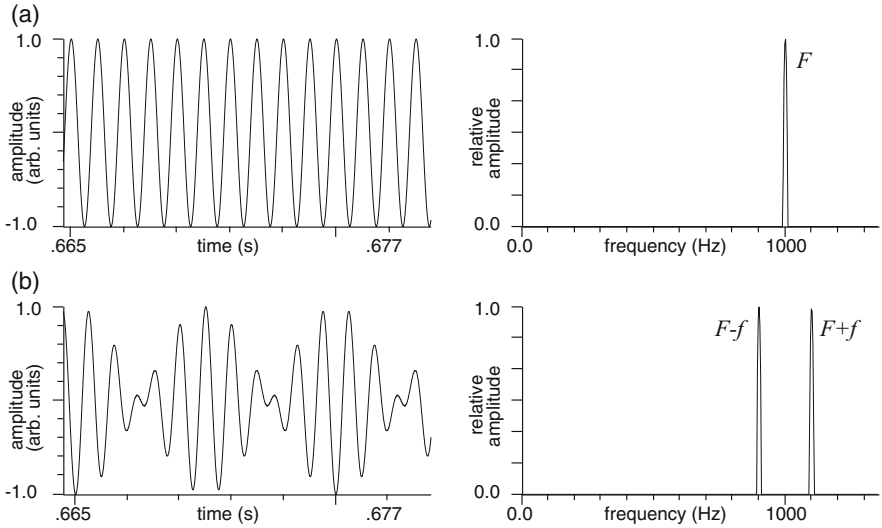

Fig. 6.8. (**a**) Harmonic signal. Its spectrum is composed of a single frequency $F = 1000\,\text{Hz}$. (**b**) The same signal as in (**a**), but with its amplitude modulated at $f = 100\,\text{Hz}$. The signature of the coupling between the two oscillations F and f (amplitude modulation, in this case) is the appearance of new frequencies: the sum and difference frequencies

be the case for the magpie tanager (*Cissopis leveriana*) [Straneck 1990b], one of its songs being displayed in Fig. 6.9. Notice that the traces in the sonogram, to the left of the dotted line, denote the existence of two independent sources: the upper and lower traces are related neither temporally nor harmonically

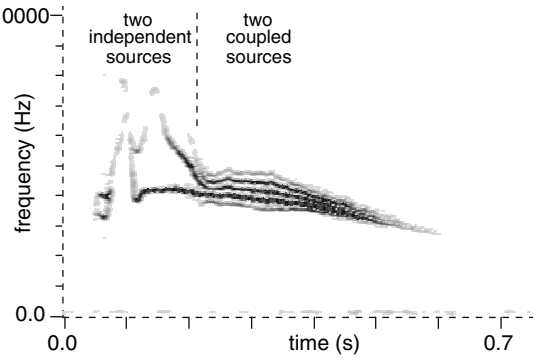

Fig. 6.9. In this syllable of a magpie tanager's song (*Cissopis leveriana*), we see at the begining the typical signature of two independent sound sources (on the two sides of the syrinx). As the frequencies become closer, a series of subharmonics can be observed. It is likely that coupling between the two sources is acting here

(i.e., the upper traces are not multiples of the lower ones). To the right of the dotted line, the situation is different: even though there is no harmonic relationship between them, there are common temporal dynamics. The bird has modified its vocal organ and/or its tract in order to make the sources interact. The result is the appearance of subharmonics, as can be observed in this complex sonogram. In this case, both complex neural instructions and complex physics are in play to generate a unique (and, in some cases, proven to be biologically relevant [Aubier et al. 2000]) sound.

6.4.2 Modeling Two Acoustically Interacting Sources

A quantitative description of this phenomenon requires us to take into account the dynamics of the two pairs of labia. The simplest model to describe the two pairs can be expressed in terms of the variables x_i and y_i $(i = 1, 2)$, describing the deviations from the rest position and the velocities of the labia in each of the sources:

$$\begin{aligned} \dot{x}_1 &= y_1 \,, \\ \dot{y}_1 &= -k_1 x_1 - b y_1 - c x_1^2 y_1 - f_{01} + p_{g1} \,, \end{aligned} \tag{6.22}$$

$$\begin{aligned} \dot{x}_2 &= y_2 \,, \\ \dot{y}_2 &= -k_2 x_2 - b y_2 - c x_2^2 y_2 - f_{02} + p_{g2} \,, \end{aligned} \tag{6.23}$$

where k_i are the restitution constants for each side, b is the dissipation constant, c is the coefficient of the nonlinear terms bounding the labial motion, and f_{0i} are force terms controlled by the activity of the dorsal muscles, all per unit mass of the labia. Finally, p_{gi} are the averaged interlabial pressures on each side.

The terms describing the pressure between the labia on each side, p_{gi}, depend on both the sublabial pressure p_s and the pressure at the input of the trachea p_i. Considering that the dynamics of each pair of labia can be described in terms of a flapping model (as discussed in Chap. 4), we can write

$$p_{gi} = p_s - \left(\frac{a_2}{a_1}\right)_i (p_s - p_i) \,, \tag{6.24}$$

where p_i stands for the air pressure at the input of the trachea, which is common to both sources. The ratios $r_i \equiv (a_2/a_1)_i$ between the glottis exit and entrance areas can be approximated by a linear function of the velocity of the ith labium, i.e., $r_i \sim -y_i$; this term is therefore directly responsible for the onset of the oscillations, since it is capable of compensating dissipative losses.

We further assume that the main constribution to the pressure fluctuations at the input of the trachea is given by the time derivatives of the airflow (inertive coupling). With this and some additional assumptions (among

which that there is no reflection at the end of the vocal tract), p_i can be approximated by

$$p_i = \beta(y_1 + y_2)\,. \tag{6.25}$$

The coupling parameter β is analogous to the acoustic inertance I in (6.11), up to a coefficient made up of a combination of anatomical parameters. Notice that the pressure at the input of the trachea p_i keeps track of the oscillations at both of the sources, and therefore couples the two oscillations.

A simulation using this model is shown in Fig. 6.10. The syllable was generated by driving (6.22)–(6.25) with a time-dependent p_s, in much the same way as the syllables of Chap. 5 were. The remaining parameters were symmetric (except the frequency-controlling parameters k_1 and k_2) and constant in time. The isolated-source spectra are apparent, with fundamental frequencies F_1 and F_2 around 2000 Hz. In addition, the signature of coupling is readily seen: the heterodyne frequencies f_{mn} appearing as light strokes in the sonogram, which are sums and differences of multiples of F_1 and F_2:

$$f_{mn} = mF_1 + nF_2\,, \quad m, n \in Z\,. \tag{6.26}$$

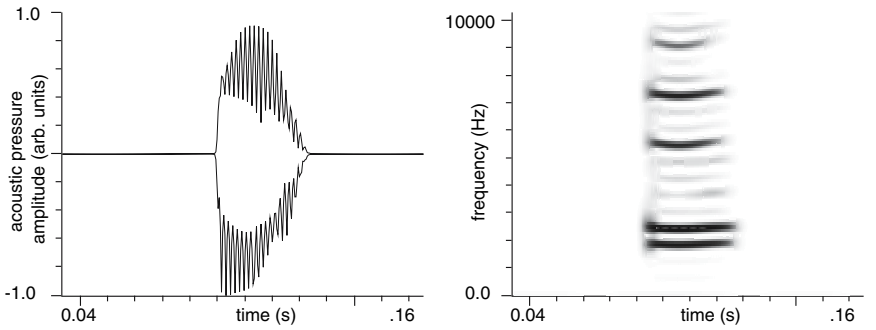

Fig. 6.10. Synthetic syllable generated with the theoretical model for coupled sources (6.22)–(6.25). The model parameters are all symmetric except for $k_1 \neq k_2$. In addition to the spectral components of the isolated sources (the two dark strokes around 2000 Hz and their corresponding harmonics), a series of lighter strokes appear. These new frequencies are sums and differences of multiples of the two fundamental frequencies

6.4.3 Interact, Don't Interact

But now the question arises of why the two sources do not couple in general. In principle, the bird has the ability to control the bilateral structure. For example, control of the dorsal muscles (gating muscles) is lateralized. Consequently, unilateral silencing is possible by active closing of one side of the syrinx.

It is natural to ask whether all the muscles involved in birdsong production reflect the lateralized activity of the vocal organ during singing. This question was settled by Goller and Suthers [Goller and Suthers 1999], who monitored bilateral airflow and subsyringeal air sac pressure in brown thrashers (*Toxostoma rufum*) during singing, together with the electromyographic activity of the abdominal muscles. They found that expiratory muscles showed activity on both sides, regardless of whether the song was produced bilaterally or on one side of the syrinx. The motor commands to the respiratory muscles appear to be bilaterally distributed, in contrast to what happens with the motor control of the syrinx. In our model, this is guaranteed by the introduction of a unique parameter p_s as the sublabial pressure.

Anatomical maneuvers at the level of the syrinx or the vocal tract, other than active lateral gating, can make the coupling parameter α change its value and vary the coupling strength as the bird sings. Coupling can then be turned on and off "on the fly". On the other hand, coupling may always be present but have little effect unless the sources are vibrating at nearby frequencies (in terms of our model, when the parameters k_i governing the frequencies on each side are very close to one another). This seems to be the case for the sonogram in Fig. 6.9.

Acoustic interaction, however, is not the only way to achieve interaction. The coupling between the sources may also be structural, for example involving the cartilaginous pessulus to which the labia on both sides of the syrinx are attached [Nowicki and Capranica 1986]. Little is known about this, but the dynamics displayed by two nonlinear, mechanically coupled oscillators cannot but be exciting.

These examples of complex sounds are extremely interesting, since the source of their complexity should not be traced up towards the brain, but to the nonlinearities of the peripheral system. Chaotic calls [Fletcher 2000], period doubling [Fee et al. 1998], etc. are only a few examples of a large class of nonlinear phenomena present in vocalizations across the animal kingdom [Wilden et al. 1998]. In recent work, Tchernikovski et al. [Tchernikovski et al. 2001] even showed that zebra finches take advantage of period-doubling bifurcations in their learning processes. These authors showed that the acoustic features of syllables would show, during development, an increasing acoustic mismatch until an abrupt correction would take place (a period-doubling bifurcation). This process might reflect the physical and neural constraints on the production and imitation of song. Along the same lines, Podos [Podos 1996] showed that syntax development could also be affected by the physical limits of how birds sing, by the introduction of pauses, stops, etc., opening a new perspective on the issue of how vocal diversity arose. Altogether, these results illustrate that the final vocal output is the result of a rich interplay between neural instructions and a nontrivial physical organ that provides limitations as well as opportunities [Goller 1998, Fee et al. 1998, Chiel and Beer 1997].

7 Synthesizing Birdsong

Understanding the mechanisms behind birdsong production should enable us to generate realistic synthetic song. A synthesizer could be built in such a way that the acoustic properties of the song were reproduced by mechanisms that were qualitatively different from those found in the actual birds, but this is not our goal. We have focused on the generation of synthetic song by reproducing the mechanisms found in the real system, as a way to build confidence in our modeling. Other applications of our synthesizers will be discussed at the end of the chapter. Two different strategies were followed. First, we have produced synthetic song by generating audio files that can be read by PC sound player software. Those audio files contain synthetically generated sound pressure time series data, obtained by numerical integration of the equations used to model the mechanisms involved in birdsong. Our second strategy consisted of building an analog integrator of those equations, using commercial integrated circuits. The output can be listened to by simply connecting a loudspeaker to the electronic circuit.

7.1 Numerical Integration and Sound

In previous chapters, we derived systems of differential equations that would allow us to compute the dynamical evolution of the variables involved. In particular, the midpoint of a labium x and its velocity y were found to obey (in a very simple model) a pair of first-order differential equations

$$\begin{aligned} \dot{x} &= y\,, \\ \dot{y} &= -\epsilon(t)\,x + (\beta(t) - b)\,y - cx^2 y\,, \end{aligned} \tag{7.1}$$

where $\epsilon(t)$ and $\beta(t)$ are slowly varying temporal functions, periodic on the timescale of the syllable, for phrases consisting of repeated syllables. The time evolution of x and y, that is, the rule $t \to (x(t), y(t))$, is in principle completely determined by the differential equations above, once we have chosen an initial condition $x(t_0) = x_0$, $y(t_0) = y_0$. But how can we actually compute the solutions $\mathbf{x}(t) = (x(t), y(t))$? Analytical formulas are out of the question, but good approximations are possible.

7.1.1 Euler's Method

The simplest possible method for integrating a differential equation is known as *Euler's method*. The basic idea behind it is the following. Given a function $\mathbf{x}(t)$, we know that a good approximation to its derivative $d\mathbf{x}/dt$ (or $\dot{\mathbf{x}}$) can be computed from

$$\dot{\mathbf{x}}(t_0) = \frac{\mathbf{x}(t_0 + \Delta t) - \mathbf{x}(t_0)}{\Delta t}\,, \tag{7.2}$$

for a small enough Δt. Conversely, given the derivative $\dot{\mathbf{x}}$ of a function $\mathbf{x}(t)$ and an "initial" value of the function $\mathbf{x}(t_0)$, we can compute an estimate of the value of the function at a time $t_0 + \Delta t$ from

$$\mathbf{x}(t_0 + \Delta t) = \mathbf{x}(t_0) + \dot{\mathbf{x}}(t_0)\,\Delta t + O((\Delta t)^2)\,, \tag{7.3}$$

and we do know $\dot{\mathbf{x}}$ here. An approximation to the solution $\mathbf{x}(t)$ for all t is achieved by concatenating steps such as this using the estimates from the step before.

This approximation is not very good, actually, unless the time step Δt is extremely small. The main problem with this method is that one computes the derivative values at the left-hand end of each time interval (t_n, t_{n+1}). Therefore, the error is only one power of Δt smaller than the correction. This can be seen by continuing the expansion of (7.3), which shows that the error is of order $O((\Delta t)^2)$.

An improvement over Euler's method is to estimate the value of the derivative at other points of the interval and then take an average in order to compute the values of the variables at time $t + \Delta t$, an algorithm which is known as the *improved Euler method*. For example, in a given step, one can compute a trial value

$$\mathbf{x}_{trial}(t_{n+1}) = \mathbf{x}(t_n) + \dot{\mathbf{x}}(t_n)\,\Delta t \tag{7.4}$$

and use it to estimate an average of the time derivative, to obtain

$$\mathbf{x}(t_{n+1}) = \mathbf{x}(t_n) + \frac{\dot{\mathbf{x}}(\mathbf{x}(t_n)) + \dot{\mathbf{x}}(\mathbf{x}_{trial}(t_n))}{2}\,\Delta t \tag{7.5}$$

(notice that we know $\dot{\mathbf{x}}$ as an explicit function of $\mathbf{x}$ through the vector field $\dot{\mathbf{x}} = f(\mathbf{x})$). This method also gives an error that goes to zero as the time step goes to zero. But, compared with Euler's method, the error of our new method tends to zero faster.

7.1.2 Runge–Kutta Methods

Another method frequently used to integrate ordinary differential equations is known as the *Runge–Kutta method*. Its simple implementation uses a trial step as before. The derivative at the begining of the interval is used to find a

point halfway across the interval. This is our trial step. Using the derivative at this estimated point, we advance the dynamics across the full width of the interval. By this procedure, we obtain second-order accuracy.

We can illustrate this method with a one-dimensional example: the integration of $\dot{x} = f(x)$. We define an estimate of the value of the variable at an instant $t + \Delta t$ by

$$x_{estimated}(t + \Delta t) = x(t) + \Delta t f\left(x + \frac{1}{2}\Delta t f(x)\right), \tag{7.6}$$

where $f(x + \frac{1}{2}\Delta t f(x))$ is $\dot{x}$ evaluated at the point $x + (1/2)\Delta t f(x)$. On the other hand, we can expand $x(t + \Delta t)$ as a Taylor series in Δt, and use it to compute the *error*,

$$error = \|x(t + \Delta t) - x_{estimated}(t + \Delta t)\|. \tag{7.7}$$

Expanding $x(t + \Delta t)$ as a Taylor series in Δt, we obtain $error = O(\Delta t^3)$, since the quadratic terms cancel out.

It is natural to ask then why we should not continue with the process of estimating the time derivatives at yet more points, and obtain in that way an approximation which is better and better for a given size of time step. There is nothing wrong with that argument... except that, at some point, what we gain on one side is lost on the other side, since many functions have to be evaluated. A good compromise is known as the fourth-order Runge–Kutta method, which, for a system of equations written as

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \tag{7.8}$$

estimates

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \frac{1}{6}(\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4), \tag{7.9}$$

where

$$\mathbf{k}_1 = \mathbf{f}(\mathbf{x}_n)\Delta t, \tag{7.10}$$

$$\mathbf{k}_2 = \mathbf{f}\left(\mathbf{x}_n + \frac{1}{2}\mathbf{k}_1\right)\Delta t, \tag{7.11}$$

$$\mathbf{k}_3 = \mathbf{f}\left(\mathbf{x}_n + \frac{1}{2}\mathbf{k}_2\right)\Delta t, \tag{7.12}$$

$$\mathbf{k}_4 = \mathbf{f}(\mathbf{x}_n + \mathbf{k}_3)\Delta t. \tag{7.13}$$

Typically, very efficient routines are used to implement these techniques. The interested reader is encouraged to consult classical references such as *Numerical Recipes in C* [Press et al. 1992], which contain widely tested subroutines that compute the quantities above. In this way, writing a code that approximates the solutions of our differential equations is as easy as implementing a loop in which, at each step, the subroutine is called.

7.1.3 Listening to Numerical Solutions

With the methods described above, we can compute a good approximation to the dynamics of the labial motion. As we have already discussed, the pressure fluctuations at the input of the trachea $p_i(t)$ are the result of two perturbations: those induced by the modulation of the airflow due to labial motion $s(x(t), y(t))$, and the pressure waves that return to the base of the trachea after being partially reflected back at the beak, p_{back}:

$$p_i(t) = s(x(t), y(t)) + p_{back}(t - L/c)\,, \tag{7.14}$$

$$p_{back}(t) = -\gamma p_i(t - L/c)\,, \tag{7.15}$$

where γ is the coefficient for internal reflection at the beak, and is equal to 1 in the ideal case of perfect reflection.

Computing x and y numerically allows us to obtain the air velocity at the open end of the tract. In order to do so, we remember that $\nabla p = -\rho_0 \partial \mathbf{v}/\partial t$, and use this relation to compute the velocity. The total pressure at a small distance dx from the end of the tube is

$$p_{tot}(x = L - dx) = p_{back}(t - dt) + p_i(t - (T - dt))\,, \tag{7.16}$$

where $dx = c\,dt$. Since, at the boundary, $p_{back}(t) = -p_i(t - T)$, we can write

$$p_{tot}(x = L - dx) = -p_i(t - T - dt) + p_i(t - (T - dt)) = 2\frac{dp_i}{dt}\,|_{t-T}\,dt\,. \tag{7.17}$$

Computing the gradient from

$$\nabla p_i = \frac{0 - 2\,(dp_i/dt)|_{t-T}\,dt}{dx} \tag{7.18}$$

and performing a temporal integration, we obtain the following for the velocity at the open end of the tube:

$$v = \frac{2}{c\rho_0} p_i(t - T)\,. \tag{7.19}$$

We can use the numerically computed time series representing the air motion to excite a mechanical membrane and generate a physical sound wave. Software players actually allow you to do this: they convert numerical instructions into electrical signals at the output of a sound card, which are capable of driving a loudspeaker. The tricky part of this procedure is to discover the format of the file that is interpreted by the player. One of these formats is called "wav". In this format, the sound is represented by a string of integers in the interval $(-32768, 32767)$, in a binary format (i.e., each value of p is represented by the coefficients b_i, where $p = \sum b_i 2^i$). In this way, we can do the following.

1. Integrate the equations of the labial motion $\mathbf{x}(t)$.

2. Use the values of $\mathbf{x}$ to generate the sound pressure fluctuations and displacements at the end of the tract.
3. Normalize these values and write them in binary format.

We are then ready to listen to our sounds. Windows operating systems are distributed with a sound player among their accessories, and a variety of sophisticated options are common. Linux is also distributed with sound players, and many distributions also come with *snd*, an open-source sound player, recorder and editor.

It should be pointed out that we are working under the source–filter theory, which states that a vibrating valve in the avian vocal organ is capable of generating a multifrequency signal, which is then filtered (see Chap. 6). Are there many bird sounds which are generated by mechanisms compatible with this hypothesis? Pure-tone birdsong, for example, is a common phenomenon in birds. Since these sounds are spectrally "pure", and sound "whistled", many hypotheses have been proposed to account for them. For example, it is possible to think of tube resonances coupling to a vibrating valve, suppressing the normal production of harmonic overtones. In order to settle this issue, Beckers and coworkers recorded sound signals close to the syringeal sound source during spontaneous pure-tone vocalizations. At least for turtledoves, these authors could show that pure tones originate through filtering of a multifrequency harmonic sound source [Beckers et al. 2003a].

7.2 Analog Integration

If we feel uncomfortable around computers, we can actually build an analog integrator of the equations defining our model, and physically listen to the synthetic songs by connecting it appropriately to a loudspeaker [Reuter et al. 1999]. In this section we describe a very simple analog synthesizer, built with commercial electronic components.

7.2.1 Operational Amplifiers: Adding and Integrating

Our main building block is the *differential amplifier*, a device with two inputs (positive and negative, or noninverting and inverting) as shown in Fig. 7.1, with an output voltage $V_{out} = A\,(V_+ - V_-)$, where A is the gain. A special form of differential amplifier, the *operational amplifier* or *op-amp*, is widely used. It has an extremely large gain (typically on the order of 10^6). In fact, in use, these devices are connected to power supplies of about $\pm 15\,\mathrm{V}$, which of course bounds their amplification gain.

As described above, it would seem that op-amps are useful only for tiny signals, but these devices are used with connecting impedances between one of the inputs and the output, which provides a negative-feedback loop. Under these conditions, the op-amp will operate in such a way that the following conditions apply:

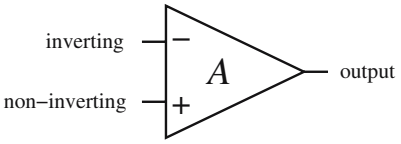

Fig. 7.1. A differential amplifier. The device has two inputs, V_+ and V_-, and an ouput $V_{out} = A\,(V_+ - V_-)$, where A is the gain of the amplifier

1. The voltages satisfy $V_+ \sim V_-$.
2. The current into the device is negligible.

It is not complex to show that in this operating regime, the device illustrated in Fig. 7.2a is a summing amplifier: since one of the inputs (V_+) is grounded, $V_- \sim 0$. Therefore, the currents through the resistances are $I_j = V_j/R_j$. Since the device takes virtually no current, the sum of all those currents passes through R_F. Therefore,

$$\frac{V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_3}{R_3} = -\frac{V_{out}}{R_F}\,. \tag{7.20}$$

By choosing the resistances, the output voltage can be chosen to be (modulo an inversion) the sum of the input voltages.

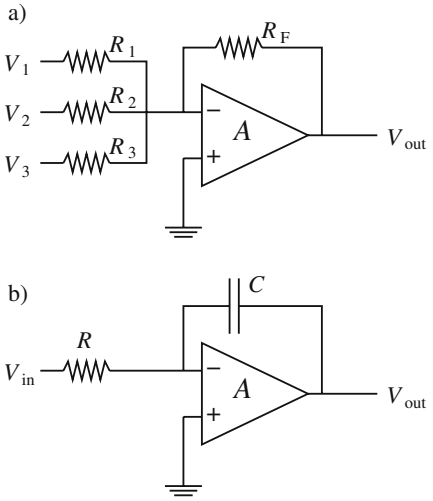

Fig. 7.2. (**a**) A summing amplifier. This arrangement of resistances and an op-amp allows us to obtain an output voltage that is the sum of a number of voltages in the circuit. (**b**) An integrator. By connecting a capacitor and a resistance as shown in the figure to an op-amp, we are able to integrate the signal V_{in}

A second important configuration of an op-amp, for the purpose of building our synthesizer, is the one displayed in Fig. 7.2b. The current charging the capacitor is simply $I = V_{in}/R$, since no current enters the op-amp. On the other hand, since $V_- \sim V_+ = 0$,

$$\frac{V_{in}}{R} = -C\,\frac{dV_{out}}{dt}\,. \tag{7.21}$$

Therefore,

$$V_{out} = -\frac{1}{RC}\int V_{in}\,dt\,, \tag{7.22}$$

allowing us to integrate the input signal. With these elements, we have built an analog integrator for the simplest model describing labial motion in the syrinx.

7.2.2 An Electronic Syrinx

In Fig. 7.3, we display an analog integrator for the system of equations under analysis. Following the circuit clockwise from the top, we see first an arrangement of an op-amp, resistor and capacitor like the one displayed in Fig. 7.2b, i.e., an integrator. The voltage obtained is $v = -(1/RC)\int w\,dt$. The voltage V_a is obtained by multiplying v and an external signal $\epsilon(t)$, which represents the varying tension of the syringealis ventralis muscle. The operation of multiplying can be performed by means of an integrated circuit (for example, an AD633JN multiplier), which multiplies the signals and divides the result by

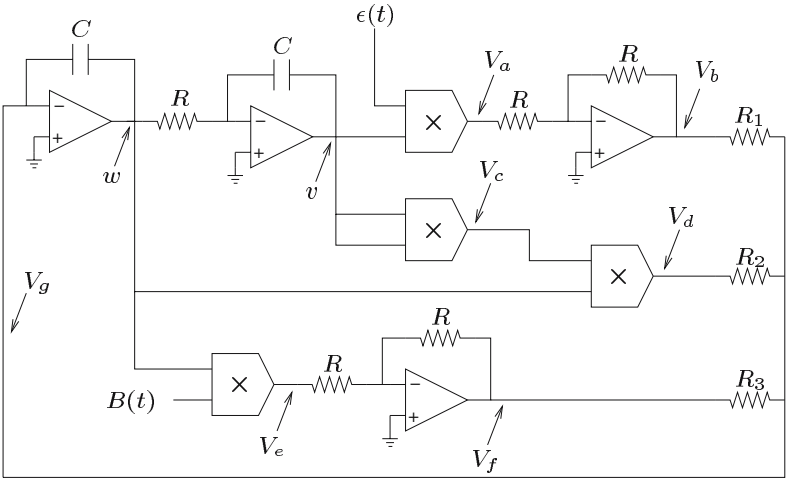

Fig. 7.3. An electronic syrinx. This device integrates by an analog method the equations of a very simple model for the labial motions. When it is connected to a loudspeaker, a synthetic song is generated

a factor of 10 V. The voltage V_b is just $-V_a$. By means of another multiplier, we construct $V_c = v^2/(10\,\mathrm{V})$, which is in turn multiplied by w in order to generate the function $V_d = w\,v^2/(100\,\mathrm{V}^2)$.

In the third layer of the circuit, the external function $\beta(t)-b$ emulating the cyclic changes in bronchial pressure above the linear dissipation is multiplied by the voltage w, generating the voltage $V_e = w\,(\beta(t) - b)/(10\,\mathrm{V})$, which, after an inverting amplifier, gives rise to $V_f = -w\,(\beta(t) - b)/(10\,\mathrm{V})$. Finally, the voltages V_b, V_d and V_f are fed into the integrator at the top left through R_1, R_2 and R_3, respectively. Therefore,

$$w = -\frac{1}{C}\int\left(-\frac{v\epsilon(t)}{R_1\times 10\,\mathrm{V}} + \frac{wv^2}{R_2\times 100\,\mathrm{V}^2} - \frac{w\,(\beta(t)-b)}{R_3\times 10\,\mathrm{V}}\right)dt\,. \tag{7.23}$$

As a result, calling $V_x = -v$ and $V_y = w/(RC)$, we obtain

$$\dot{V}_x = V_y\,, \tag{7.24}$$

$$\dot{V}_y = -\frac{\epsilon(t)}{RR_1C^2\times 10\,\mathrm{V}}\,V_x - \frac{1}{R_2C\times 100\,\mathrm{V}^2}\,V_x^2V_y + \frac{\beta(t)-b}{R_3C\times 10\mathrm{V}}\,V_y\,, \tag{7.25}$$

which have the same form as (7.1), our simple model. We chose $C = 0.1\,\mu\mathrm{F}$, $R_1 = 10\,\Omega$, $R_2 = 33\,\Omega$, $R_3 = 10\,\mathrm{k}\Omega$ and $R = 10\,\mathrm{k}\Omega$.

The integrator that we have described can be driven with our forcing functions $\beta(t)$ and $\epsilon(t)$, which emulate the time-varying air sac pressure and the tension of the ventral muscles, respectively. In the implementation that we are discussing, the amplitudes of the forcing functions can be in the order of volts, and the frequencies in the range 1–10 Hz. For small values of the function $\beta(t)$, the voltage v basically follows the forcing, but beyond a critical value of $\beta(t)$ the voltage v displays large-amplitude oscillations, at a frequency much larger than the forcing frequency. This activity represents the oscillation of the labia in our problem. At this point, we say, in terms of the bifurcation diagrams in Chap. 5, that the values of $\beta(t)$ take the system into the region of oscillations. For the values of capacitances and resistances chosen, these rapid oscillations have frequencies in the order of kHz.

How can we visualize our integration? In a lab, a simple oscilloscope can be used. But for a simple application such as this one, and for the ranges of frequencies involved, the data can be downloaded to a PC by sending v to the input of a commercial loudspeaker, the output of which is sent to a commercial PC sound card, and the signal can be recorded by any sound software. The free software *snd*, running under a Linux operating system, and several others programs running under Windows, are good options for analyzing the voltage time series obtained from our circuit, as well as for processing the signals and computing their spectral properties.

In Fig. 7.4, we show the sonogram of a "composition" generated by the device. In order to generate this audio file, four different recordings were made. Different values of the frequency of the forcing parameters $\beta(t)$ and

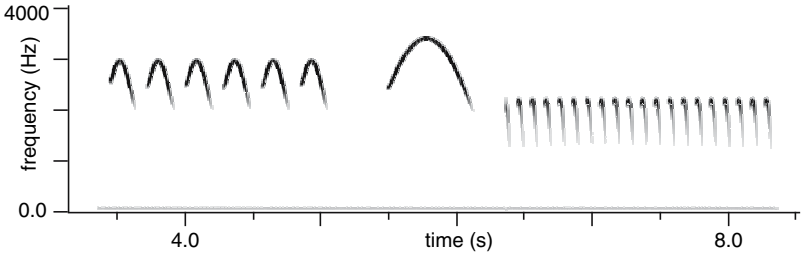

Fig. 7.4. A "composition" interpreted by the electronic syrinx. Each phrase is a repetition of a syllable. In order to generate different syllables, the frequency of the basic perturbation was changed, as well as the phase difference between the signals emulating pressure and tension

$\epsilon(t)$ were used, and the phase differences between them were varied for the different syllable types.

In order to generate different phase differences between $\beta(t)$ and $\epsilon(t)$, a sinusoidal function generated by a function generator can be amplified by means of different amplifiers such as the ones displayed in Fig. 7.5 to produce $\beta(t)$ and $\epsilon(t)$. In this subsection we have focused on the generation of signals corresponding to the labial oscillations in the syrinx, disregarding the filtering that occurs in the tract. Notice that for some species, this can be a good approximation to the real sound (as discussed in Chap. 3), although not for all.

The first inset of Fig. 7.5 displays the amplifier used to generate the instruction $\epsilon(t)$, emulating the gesture corresponding to changes in the tension of the ventral syringeal muscle. By decreasing the variable resistance, the amplitude of the cyclic gesture was increased. This caused the synthetic sounds to sweep a large range of frequencies. The potentiometer controls the constant term in $\epsilon(t)$, and therefore the average frequency of the vocalization.

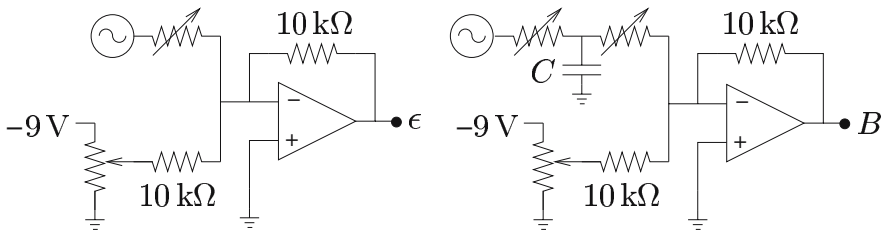

Fig. 7.5. These amplifiers can be used to generate two periodic functions with a variable phase difference. An interesting way to drive this "electronic syrinx" would be with a voltage measured by a pressure transducer connected to a cannula inserted in an air sac, and a voltage obtained after rectifying and integrating a voltage measured by microelectrodes inserted in the ventral muscles controlling the frequency of the vocalization

The cyclic gestures giving rise to syllables can be generated with a second oscillator with the same design as the one used to emulate the labial motion, but with $R_1 \approx 2\,\mathrm{M}\Omega$. This allows the device to run on 9 V batteries, and avoids the use of a function generator.

7.3 Playback Experiments

One of the purposes of synthesizing song is to build confidence in our models. If, by prescribing mathematical rules for the dynamics of the avian vocal organ, we can emulate a behavior leading to realistic sounds, we build confidence in our having identified the appropriate physical mechanisms underlying vocal production in birds. Beyond this satisfaction, there are other interesting possibilities for these techniques.

A complete book could be written on the communication between birds. We are focusing in this book on the processes that they use to vocalize, but what do they "say"? There is a wide consensus on birdsong being important in at least two different behaviors: territorial defense and mating. Field scientists test their ideas concerning these issues with a variety of techniques, one of them being playback experiments. For example, a recorded song may be used to persuade a bird that an intruder is in its territory. The use of either numerical or analog simulations instead of recordings widens the range of acoustic features that could be used in playback experiments, and provides the experimenter with the possibility of experimenting with sounds well beyond what can be achieved by simply editing recordings.

Simulations can definitely play a role in experiments on learning and auditory feedback. A computer connected to a loudspeaker can easily play back a huge number of qualitatively different songs according to a prespecified rule, throughout the learning period of a newborn songbird. There is no need for previous song recording, nor subsequent classification and storage. Furthermore, with an electronic syrinx capable of analog simulation, the experimenter now has the possibility of changing detailed features of the simulated song in *real time*. Auditory-feedback experiments involving altered syllable duration, altered frequency range and altered frequency evolution (even with a downsweep changed into an upsweep) are available.

7.4 Why Numerical Work?

Most of the theoretical results discussed in the previous chapters were obtained by numerical simulations of the equations describing the physical mechanisms involved. In the previous section, we showed how to proceed by generating of synthetic data. Is it always necesary to rely on numerical tools to advance our understanding of the generation of birdsong?

In general, numerical tools are needed as soon as nonlinear effects are taken into account. Acoustics (in the regimes that we have been exploring) is a linear theory and, therefore, complex problems can be expressed in terms of simple problems. For example, the pressure waves established in a tube which is forced by a complicated excitation at one end can be computed by adding the pressure waves established when the tube is forced by a sum of simple (harmonic) excitations at the end. Whenever the physical system under study is linear, this procedure can be followed. In the problem of sound generation by an oscillating device coupled to a tube, this strategy can be followed as long as the whole system behaves linearly. For these cases, a most useful concept can be defined: *impedance*.

7.4.1 Definition of Impedance

At the core of the concept of impedance lies the following feature of forced linear oscillators: a linear oscillator forced by a harmonic function of time will ultimately end up following the driver at the forcing frequency. However, there will be phase lags betwen the forcing and the forced system. Also, the amplitude of the response will depend on how similar the forcing frequency and the natural frequency of the oscillating system are. But after a transient, the forced oscillator will follow the forcing.

As an example, let us consider a simple harmonic oscillator, whose displacement x from equilibrium is governed by the equation

$$m\frac{d^2x}{dt^2} + b\frac{dx}{dt} + kx = f(t)\,, \tag{7.26}$$

whenever it is forced by a driving force $f(t)$. An arbitrary forcing $f(t)$ can be decomposed in terms of harmonic functions as

$$f(t) = \sum F_n e^{i\omega_n t}\,, \tag{7.27}$$

as discussed in Chap. 1. Since the system under study is linear, a sum of solutions is also a solution. Therefore, the building block for understanding the general problem (7.26) is the solution of

$$m\frac{d^2\mathbf{x}}{dt^2} + b\frac{d\mathbf{x}}{dt} + k\mathbf{x} = Fe^{i\omega t}\,, \tag{7.28}$$

where the real part of the complex forcing is $\mathrm{Re}(Fe^{i\omega t}) = F\cos(\omega t)$, and therefore $\mathbf{x}$ is a complex variable whose real part represents the physical displacement of the oscillator (in this section, boldface fonts are used to indicate complex numbers). Since we know that, ultimately, the forced system will oscillate at the same frequency as the forcing, we can propose as a solution $\mathbf{x} = \mathbf{A}e^{i\omega t}$. The advantage of using these complex numbers is that taking a temporal derivative is now trivial (it is simply a multiplication by $i\omega$), and therefore we can write

$$-\omega^2 m\mathbf{A} + i\mathbf{A}\omega b + \mathbf{A}k = F\,. \tag{7.29}$$

This equation allows us to compute $\mathbf{x}$ and $\mathbf{v} = d\mathbf{x}/dt$, since it determines the complex amplitude $\mathbf{A}$ of the displacement. Notice that this gets us close to computing the only unknown quantities of the problem: how strong the reaction to the forcing will be (given by the modulus of this complex number), and the relative phase between the forcing and the displacement (given by the phase of this number). It is customary to represent this information by computing the ratio between the complex forcing and the complex velocity of the system,

$$\mathbf{Z} = \frac{\mathbf{f}}{\mathbf{v}}\,, \tag{7.30}$$

where $\mathbf{Z}$ is called the *impedance* [Kinsler et al. 1982]. In general, it is a ratio between a complex forcing and the resultant complex speed of the system, at the point where the force is applied. Therefore, this number allows us to compute the velocity with which the forced system will react to a forcing. This provides a general framework for the definition of the specific acoustic impedance presented in Chap. 6.

7.4.2 Impedance of a Pipe

In acoustics, it is natural to use this concept to take account of situations in which the geometry of the boundaries confines the pressure wave to a limited region of space. If a pressure wave is established in a pipe of cross section S, for example, we can write

$$\mathbf{p} = \mathbf{A}e^{i(\omega t + k(L-x))} + \mathbf{B}e^{i\omega t - k(L-x)}\,, \tag{7.31}$$

where we have assumed that the diameter of the pipe is much smaller than the sound wavelength. This allows us to approximate the wave as a planar wave. Using the relations described in Chap. 1 to relate the pressure to the velocity of the air,

$$\mathbf{v} = -\frac{1}{\rho_0}\int \frac{\partial p}{\partial x}dt\,, \tag{7.32}$$

we can compute the impedance of the system. At the end $x = L$, we find that

$$\mathbf{Z}_L = \rho_0 cS\frac{\mathbf{A}+\mathbf{B}}{\mathbf{A}-\mathbf{B}}\,, \tag{7.33}$$

while at $x = 0$, we find that $\mathbf{Z}_0 = S\mathbf{p}/\mathbf{u}$ is given by

$$\mathbf{Z}_0 = \rho_0 cS\frac{\mathbf{A}e^{ikL} + \mathbf{B}e^{-ikL}}{\mathbf{A}e^{ikL} - \mathbf{B}e^{-ikL}}\,. \tag{7.34}$$

Following [Kinsler et al. 1982], we combine these expressions to eliminate the amplitudes $\mathbf{A}$ and $\mathbf{B}$, writing

$$\frac{\mathbf{Z}_0}{\rho_0 c S} = \frac{\mathbf{Z}_L/\rho_0 c S + i\tan(kL)}{1 + i(\mathbf{Z_L}/\rho_0 c S)\tan(kL)}\,. \tag{7.35}$$

In order to compute the input impedance of the system (i.e., the impedance at $x = 0$), we have to compute (or estimate) the impedance at $x = L$. If we are interested in a tube *closed* at $x = L$, this is really easy. At the closed end there is no air displacement, regardless of the pressure. Therefore, the output impedance is infinite, and we obtain

$$\mathbf{Z}_0 = -i\rho_0 c S \cot(kL)\,. \tag{7.36}$$

Notice that the zeros of the impedance mean that, for a forcing with a bounded amplitude, the velocity of the system can diverge. Therefore, we can interpret these zeros as the resonances of the system.

This is a beautiful example of the power of linear methods . . . when they are valid. We can obtain analytic expressions for the conditions for resonance in terms of the system's parameters, allowing us to understand general relationships between geometry and acoustics. The problem of the open tube can be approached similarly. Unfortunately, the condition $\mathbf{Z}_L = 0$ is an oversimplification. The output impedance will be some radiation impedance (we need some force to move the air outside the tube), which is not so easy to compute [Kinsler et al. 1982].

In acoustics, we can define the *acoustic impedance* $\mathbf{z}$ of a fluid that is acting on a surface area A as the acoustic pressure divided by the volume velocity of the fluid (which is the product of the area and the particle velocity). This quantity is useful for discussing the properties of the transmission of pressure fluctuations through pipes of various geometries. The reason is that it allows us to find the impedance of a complex arrangement of tubes in terms of the impedances of the individual tubes.

Let us consider the example illustrated in Fig. 7.6, where we show a pipe that branches into two pipes, denoted as pipes 1 and 2. Let us assume that the branching pipes have acoustic impedances z_1 and z_2, and that the pressures at the junction are $\mathbf{P}_1$ and $\mathbf{P}_2$. Finally, let us denote by $\mathbf{P}_0$ the pressure at the bifurcating branch. Continuity of the pressure allows us to write

$$P_0 = P_1 = P_2\,, \tag{7.37}$$

while the continuity of the volume velocity allows us to write

$$\mathbf{U}_0 = \mathbf{U}_1 + \mathbf{U}_2\,, \tag{7.38}$$

which, combined, allows us to write that

$$\frac{1}{\mathbf{z}_0} = \frac{1}{\mathbf{z}_1} + \frac{1}{\mathbf{z}_2}\,. \tag{7.39}$$

A very nice discussion of the impedances of avian vocal organs is presented in [Fletcher and Tarnopolsky 1999]. The vocal-tract input impedances are

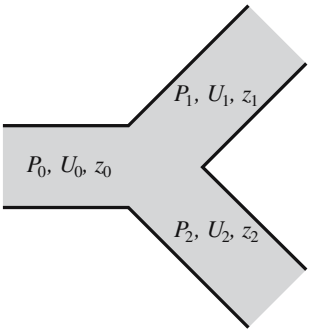

Fig. 7.6. The acoustic impedance of a branching pipe

computed as a function of the frequency, and the impedances of beaks are also discussed. A nice study of vocal-tract function in birdsong production, with manipulation of beak movements, was presented by Hoese et al. [Hoese et al. 2003].

It is possible to solve a source model to generate a harmonically rich sound signal. The volume velocity can be computed, and the component of the sound pressure associated with each harmonic component of the volume velocity can be easily obtained by just multiplying it by the proper impedance. However, a word of caution is needed. In this procedure we have assumed that the system under study is linear (the passive acoustic tract), and that the nonlinear system is uncoupled from our linear system. We are not concerned with how the spectrally rich signal that is injected into our linear problem is generated. But we must keep in mind that as soon as the tube and the nonlinear sound-generating sources are *coupled*, the system as a whole is nonlinear. At this point, direct integration of the source dynamics plus boundary conditions is needed, since the whole concept of impedance relies on the assumption that the forcing and the forced system will end up oscillating with the same frequency.

With these building blocks, some very interesting problems can be addressed. Recently, Fletcher et al. [Fletcher et al. 2004] analyzed the "coo" vocalizations of ring doves. These birds generate sounds through a mechanism that differs from those described so far. They produce a vocalization that consists basically of a pure tone of relatively low frequency (around 600 Hz). The most remarkable aspect of this vocalization is that it is produced with the beak and nostrils closed. While the bird produces this vocalization, it inflates the upper part of the esophagus, and it is this "sac" what radiates the sound. The use of impedances allowed Fletcher et al. to settle several issues. They established the plausibility of the trachea–glottis–inflated-esophagus system being a band-pass filter, and they showed that the inflated esophagus is a more efficient sound-radiating mechanism than the open beak.

8 From the Syrinx to the Brain

In the last few years there has been enormous progress in the understanding of the brain structure of birds. At least, this is the case for songbirds, which constitute some 4000 of the approximately 10 000 species of birds known to exist. The reason for such interest is that, for these species, the learning of song has many parallels with the acquisition of speech by humans [Doupe and Kuhl 1999]. Notice that we write "speech" and not "language", since there is no evidence of birds using their songs in a combinatorial way in order to convey flexible syntactic content like that characterizing human speech. However, songbirds require (as do humans) a tutor whose song is memorized in the early stages of life [Thorpe 1961, Marler 1970, Nottebohm 1970]. Later, they try to reproduce the songs previously memorized, correcting their vocalizations thanks to an auditory feedback [Konishi 1965]. For this reason, songbirds have been used as a test bench on which fundamental aspects of a complex learned behavior can be studied. This has led to the accumulation of an important amount of information on the physical substrate of these operations. This will allow us to advance further in our discussion, establishing the place in the brain where the precise gestures (responsible for the execution of the commands driving the vocal organ in order to produce a song) are controlled.

The fundamental operations involved in song learning (memorization of the song of the tutor, auditory feedback from different attempts leading to corrections, etc.) are interesting and sophisticated. For this reason, detailed experiments have been performed in order to determine whether there were specific neural circuits that would serve as the physical substrate. The first step consisted in describing morphologically the brains of songbirds. From these studies, we learned that there are sets of neurons (called nuclei) clearly grouped into different regions in the brain, interconnected in a very complex way through long axons. The way in which knowledge was generated about the role played by these nuclei was basically through the study of behavioral changes after localized lesions. As much as they help us, however, classifications can be dangerous. We shall say in this chapter that two major pathways organize these nuclei, running the risk of caricaturing a most complex biological device. In any case, it is well established in the field that a set of nuclei is necessary to generate song. Systematic studies on lesioned birds indicate

that without these nuclei, the birds cannot vocalize. For this reason, these nuclei are said to be part of the *motor pathway*. Another set of nuclei constitute the *anterior forebrain pathway*. These nuclei can be lesioned without an immediate degradation of the song. However, lesions of these nuclei affect the bird's ability to learn or maintain the vocalizations. Let us describe some features of these pathways.

8.1 The Motor Pathway

It was through lesions and observation of behavior that the nuclei involved in the generation of motor activities responsible for the production of song were identified [Nottebohm et al. 1976]. The "motor pathway" is constituted by two nuclei, called HVC and RA (robustus nucleus of the archistriatum). HVC sends instructions to RA, which in turn sends instructions to several nuclei: the tracheosyringeal motor nucleus nXIIts, which innervates the syringeal muscles, and a set of respiratory nuclei (among which we find the expiratory premotor nucleus retroambigualis RAm), which innervates the muscles involved in respiration (see Fig. 8.1). HVC and RA are necessary for song production. Lesions in them lead to the suppression of song. The specific function of each nucleus is more difficult to unveil. However, there is evidence of a hierarchical organization in the brain. Microstimulation of HVC is followed by interruption of singing and restarting of song, while microstimulation of RA disrupts only the syllable being sung [Yu and Margoliash 1996].

Observed from the peripheral system, the motor pathway has to generate cyclic instructions to drive the syrinx and the respiratory system. Moreover, in order to generate the diversity of syllables within a song, the motor pathway has to be capable of generating a diversity of phase differences between these oscillations, as discussed in Chap. 5. How does it do this? First of all, RAm and nXIIts, the brain stem structures most directly involved in phonation since they are the nuclei involved in locking expiratory activity and controlling the syrinx, are driven by projecting neurons in different regions of RA [Spiro et al. 1999]. Since RAm also projects onto nXIIts, the phase difference between the respiratory and syringeal gestures will depend on *both* the nature of this connection and the pattern of activity within RA.

The RA nucleus is innervated by excitatory projection neurons from the HVC nucleus. Just as RA is closely tied to the peripheral motor system, HVC is supposed to have connections with higher-level nervous-system functions [Yu and Margoliash 1996, Suthers and Margoliash 2002]. A song is organized in terms of minimal continuous utterances that we call syllables, combined with quiet periods, to build motifs. It has been proposed that this hierarchy in the song has a neural correlation, on the basis of the observation that electrical stimulation of HVC causes a resetting of the motif being sung, while stimulation of RA causes a distortion of the vocalization [Yu and Margoliash 1996].

This suggests that the HVC activity is sequenced to orchestrate the production of syllables and motifs, so that interrupting its progress during such a sequence causes a reset or reinitiation of the activity. It also suggests that RA is a "junction box" where HVC signals are combined to produce song, but does not itself initiate or command the song. The connections in the "junction box" can be altered by modification of its input connections from HVC and the anterior forebrain pathway, so song production can be influenced in RA but not started there.

Work on zebra finches (*Taeniopygia guttata*) has analyzed in detail the exact time relation between the firing of neurons in HVC and RA during song [Hahnloser et al. 2002]. This work has shown that during the vocalization of a motif, lasting approximately 1 s, a given RA-projecting neuron in HVC is active only during one window of temporal size 6.1 ± 2 ms in this motif, and produces 4.5 ± 2 spikes in that temporal window. If a motif is repeated, each HVC neuron will repeat its spiking in the same time window. On the other hand, RA neurons generate highly stereotyped sequences of action potential bursts, typically a few bursts of approximately 10 ms duration per motif, well time-locked with parts of syllables [Chi and Margoliash 2001]. The picture emerging from these experiments is that during each time window in the RA sequence, RA neurons are driven by a subpopulation of RA-projecting HVC neurons which are active only during that window of time [Hahnloser et al. 2002]. These experiments suggest that the premotor burst patterns in RA are basically driven by the activities of HVC neurons [McCasland 1987]. In that case, the architecture of the connectivity between the HVC nucleus and the RA nucleus will determine the complex patterns of activity in the RA neurons. Additional nuclei afferent to HVC are likely to be part of the motor pathway. One of these, the nucleus interfacialis (Nif), seems to be critical at the moment of generation of the basic oscillations which eventually drive HVC into generating the sparse activity that ultimately drives the syrinx.

8.2 The AFP Pathway

A second set of interconnected nuclei is the anterior forebrain pathway (AFP), also shown in Fig. 8.1. It connects indirectly the nuclei HVC and RA. In contrast to the motor pathway, the AFP contributes only minimally to the production of song in adults [Brainard and Doupe 2000]. However, it has been shown that lesions to these nuclei during learning profoundly alter the bird's capability for developing normal song [Bottjer et al. 1984, Scharff and Nottebohm 1991]. Three nuclei are part of the AFP: area X, the medial nucleus of the dorsolateral thalamus (DLM) and the lateral magnocellular nucleus of the anterior neostratum (lMAN). The way in which this system operates during learning is not known. The motor pathway and the AFP are not isolated. Some neurons in RA receive signals from both HVC

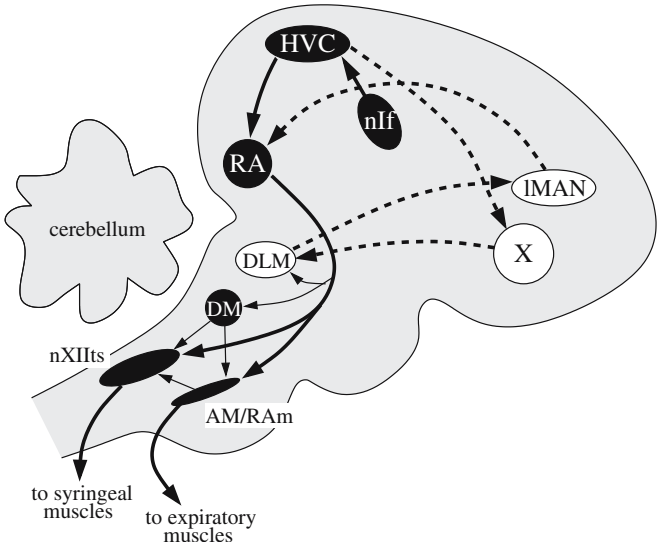

Fig. 8.1. Schematic sagittal view of the brain. The areas shaded black are necessary for adult song production, whereas the white areas are necessary for normal song development, but not adult song production. The *solid arrows* show the motor control pathway, while the *dashed arrows* show the anterior forebrain pathway, or AFP. Modified after [Spiro et al. 1999]

and lMAN. lMAN sends to RA (part of the motor pathway) either the error correction or the reinforcement signal during learning that will eventually contribute to the reconfiguration of the connections within the motor pathway. It is known that both the AFP and its connections to RA undergo regressive changes as the sensitive period ends (the sensitive period being the time in which the bird "learns" what it will eventually imitate), but the global picture is still incomplete.

8.3 Models for the Motor Pathway: What for?

When we described the physical processes involved in the generation of birdsong, we wrote down models based on first principles of physics. Can one model in neuroscience? What for? There are two main reasons to attempt to model the electrical activity of a brain when it is controlling singing. The first one is conceptual in nature. In the process of writing down a quantitative model for the voltages to be measured in the neurons involved, we realize how complete or incomplete our conceptual models are. We can, for example, write down differential equations for the voltages using the celebrated Hodgkin–Huxley equations, and test whether, for the measured neural conductances, the system generates patterns of activity like the ones measured.

In that respect, modeling can support conceptual models developed from direct experiments on one or a few neurons. But beyond that, neurons are highly nonlinear devices, and whatever smart conceptual model might be developed from experiments, it is very hard (in fact, impossible) to predict the dynamical responses that the nuclei will display by just using "intuition". Quantitative modeling can translate conceptual models into mathematical prescriptions with explicit dependences on parameters that can help turn conjectures into testable theories. On the other hand, modeling brains or nuclei remains a formidable challenge today. It is not only that computers are slow for the task. Maybe more important, there is no qualitative theory of extended nonlinear systems that could allow us to interpret the simulated information. In any case, while we wait for the architects to arrive, we shall describe the bricks.

Any modeling of brain activity will probably begin by writing equations for the time evolution of the membrane voltages of the neurons involved. The voltage is a variable that controls several biophysically relevant processes at the cell level, and is easy to measure [Koch 1999]. Just as an illustration, we shall write down the equations describing the dynamics of a simple arrangement of excitable units capable of displaying the activity in the motor pathway that we described above. This arrangement falls short of being a model of HVC since no realistic architecture or number of neurons will be included, but its building blocks have been chosen in such a way that the neural responses lie within realistic ranges. We also think that the level of description is a good compromise between simplicity and the parametric realism that could eventually help in real modeling efforts.

8.3.1 Building Blocks for Modeling Brain Activity

The majority of the cells in the nuclei that we have described in Sect. 8.2 generate brief voltage pulses, or *spikes*. These are generated in or close to the cell body, and propagate along the axon with little deformation. The ionic mechanisms that originate the spikes were first studied by Hodgkin and Huxley in the squid giant axon [Hodgkin and Huxley 1952]. These were translated into celebrated phenomenological equations that are still used to model spike generation in the nerve cells of both invertebrates and vertebrates.

The model basically assumes that the cell membrane can be characterized by a capacitance C_M due to its capacity to hold ionic imbalances with respect to the extracellular space. However, the difference of potential V between the interior and the exterior of the cell can evolve in time as ions cross the membrane through specific channels. Since the total membrane current is the sum of ionic and capacitive currents [Hodgkin and Huxley 1952] and must be zero because there cannot be a net buildup of charge on either side of the membrane, then

$$I_{ionic} + C_M \frac{dV}{dt} = 0\,. \tag{8.1}$$

The difficult part in this approach is the titanic work needed to unveil the nature of the currents I_{ionic}. Hodgkin and Huxley were capable precisely of writing down phenomenological models for these.

The two major voltage-dependent ionic currents involved in the generation of spikes are a sodium current and a potassium current. These are characterized by conductances G_{Na} and G_K, respectively. A smaller "leak" current, characterized by G_L, is also important in the description. Therefore,

$$I_{ionic} = I_{Na} + I_K + I_L \,, \tag{8.2}$$

where

$$I_i = (V - E_i)G_i(V,t) \,. \tag{8.3}$$

Here E_i are reversal potentials, typical of each of the ionic species (i = Na, K, L). The conductances $G_i(V,t)$ are typically expressed as products of maximum values g_i and nonlinear functions of coefficients that describe macroscopically the fraction of open ionic channels.

The equations end up being quite intimidating, but the dynamics displayed by such a system are not complex to describe, at least for one unit. The most remarkable property of the axonal membrane is its capacity to respond in two qualitatively different ways to depolarizing perturbations. What could be the origin of these perturbations? In its natural environment, the origin is the currents that occur when neurotransmitters are released by another neuron in a synapse, and ionic channels are opened. The qualitatively different responses can be described as either (a) a small depolarization followed by a return to the resting potential, or (b) a pulse-like action potential, whose shape is somewhat independent of the perturbation. The last response will occur whenever the perturbation exceeds a threshold.

In order to communicate, neurons have specialized contact zones called *synapses*. In what is known as a *chemical* synapse, a spiking presynaptic neuron releases *neurotransmitters*, chemicals which are capable of opening channels that allow the passage of ions through the membrane of a postsynaptic neuron. When the synapse is an excitatory one, the postsynaptic membrane potential rapidly depolarizes, finally to return to its rest value. When the synapse is inhibitory, a hyperpolarization takes place. Therefore, the result of spiking activity in a presynaptic neuron is reflected in the onset of a postsynaptic current

$$I_{syn} = (V(t) - E_{syn})G_{syn}(V_{pre}, t) \,, \tag{8.4}$$

where $G_{syn}(V_{pre}, t)$ is determined by the presynaptic state. The details of the function $G_{syn}(V_{pre}, t)$ depend on the nature of the synapse. Most of the fast excitatory synapses are regulated by a neurotransmitter called glutamate, and various receptors are sensitive to it (with different synaptic properties associated with different receptors). The inhibitory synapses are regulated by γ-amino butyric acid (GABA), the most common inhibitory neurotransmitter

in invertebrates and vertebrates. In order to take into account the effect of the synapse in our model, a current such as that in (8.4) should be added to (8.1).

The details of these models can be found in a vast literature on computational neuroscience [Koch 1999]. We would like to point out at this point that the simplicity of the description of the dynamics that one neuron can display (its *excitable* nature) can be misleading. Once several units are coupled, a wide variety of complex dynamics can be found. We can couple bidirectionally an excitatory neuron with an inhibitory one, and oscillations can take place [Hoppensteadt and Izhikevich 1997]. We can couple a simple neural oscillator to a third neuron, and a variety of subharmonic dynamics can be found [Feingold et al. 1988, Sigman and Mindlin 2000]. It is not our intention to intimidate the reader with the complexity that can emerge out of coupling excitable systems, but just to stress that endless complex labyrinths can be built with these simple building blocks.

8.4 Conceptual Models and Computational Models

It is experimentally well supported that the RA-projecting neurons of HVC burst sparsely during a motif. These units will then recruit neurons in RA, following an architecture of connections which will, in part, be responsible for the acoustic features of the song produced [Abarbanel et al. 2004a, Fee et al. 2004].

In Fig. 8.2, we display a schematic picture of this conceptual model. A set of excitatory and inhibitory units represents a subpopulation of HVC neurons, some of which project to the nucleus RA. This nucleus is represented by a set of excitatory and inhibitory units, the excitatory ones with local connections and the inhibitory ones with long-range connections.

One can also translate this picture into a computational model, in which the equations described in the previous section are computationally implemented in order to emulate the dynamics of the variables describing the activities in the nuclei. In Fig. 8.3, we show an example. The time evolution of the voltages in two RA-projecting HVC neurons is displayed in Fig. 8.3a. These two excitatory units were connected to an inhibitory one. Unless externally forced, these excitatory units did not fire. The connections between these two units were such that the second unit would spike if the first one did. Then, both units were connected to one neuron in RA, and the resulting activity is displayed in Fig. 8.3b. This architecture allows us to reproduce experimentally observed data, in which sparse spiking activity in HVC recruits neurons in RA.

Reproducing an experimental result is an important first step in a model, since it builds confidence in the plausibility of the proposed mechanisms. However, models exhibit their usefulness when they allow us to explore new regimes not yet experimentally observed. A simulation on a larger scale could

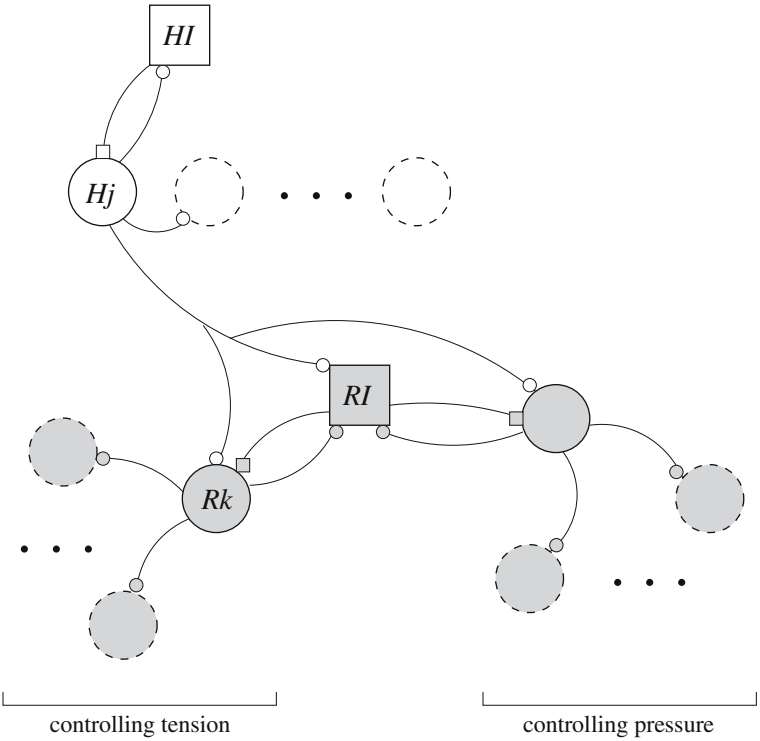

Fig. 8.2. A schematic illustration of the connection between excitable units in the nuclei HVC and RA. Inhibitory neurons are represented by squares, and excitatory neurons are represented by *circles*. HVC units are *white*, and RA units are *shaded gray*. *Square* and *round* connections represent inhibitory and excitatory connections, respectively. For the sake of clarity, only a few representative units and connections are shown

be attempted, using the architecture shown in Fig. 8.2. What equations allow us to "translate this picture into a computational model"?

8.4.1 Simulating the Activity of HVC Neurons

In order to generate Fig. 8.3, we integrated conductance-based equations for the membrane voltages of HVC and RA neurons. In this and the following subsections, we write these equations down. The arrangements of neurons used to emulate the behavior of the nuclei are much simplified. However, this simplification will give us an insight into the way to proceed in order to carry out this kind of simulation.

We write the equations for a set of N_{HVC} HVC $\rightarrow$ RA projection neurons, denoting the time-dependent membrane potential for the jth HVC excitatory neuron by V_{Hj}, where $j = 1, 2, \ldots, N_{HVC}$. This potential satisfies the

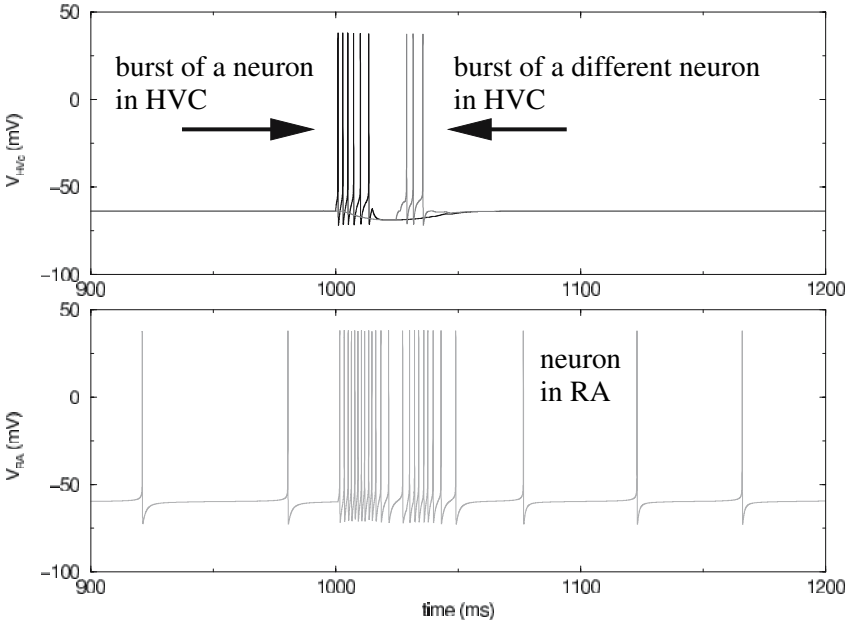

Fig. 8.3. The voltage of two units emulating neurons in HVC (**a**), and the voltage of a unit emulating a neuron in RA (**b**). Adapted from [Abarbanel et al. 2004a]

Hodgkin–Huxley (HH) equation

$$C_M \frac{dV_{Hj}}{dt} = I_L^j + I_{Na}^j + I_K^j + I_{HE}^j + I_{HI}^j \,, \tag{8.5}$$

where we have included the leak, sodium and potassium currents in the fashion described in the previous section. In addition, a synaptic current I_{HE} from the previous excitatory unit is considered, as well as a synaptic current I_{HI} from the single inhibitory unit. When isolated, each excitatory HVC neuron is silent and has a resting potential of approximately −65 mV. Leak and ionic currents are constructed, as before:

$$I_L^j = (E_L - V_{Hj}) G_L^j \,, \tag{8.6}$$

$$I_{Na}^j = (E_{Na} - V_{Hj}) G_{Na}^j \,, \tag{8.7}$$

$$I_K^j = (E_K - V_{Hj}) G_K^j \,, \tag{8.8}$$

where E_{Na} and E_K are the equilibrium potentials of the corresponding ionic currents, and E_L is the equilibrium potential for the leakage current. G_L^j is the (constant) conductance for the leak current, and G_{Na}^j and G_K^j are the voltage-dependent, ion-specific channel conductances, which are given by

$$G_{Na}^j(V_{Hj}) = g_{Na} m_j^3 h_j \,, \tag{8.9}$$

$$G_K^j(V_{Hj}) = g_K n_j^4 \,, \tag{8.10}$$

where g_{Na} and g_K are the (constant) maximum conductances. The *sodium* and *potassium activation* and *sodium inactivation* variables m_j, n_j and h_j, respectively, satisfy the usual first-order kinetics:

$$\frac{dx_j}{dt} = (1 - x_j)\alpha_x - x_j\beta_x \,, \tag{8.11}$$

where x stands for m, h or n. The reader may be surprised at the exotic dependence on m^3h and n^4. These variables were phenomenologically introduced by Hodgkin and Huxley (a discovery that was worth a Nobel Prize ten years after its publication), and only later were associated with the fundamental subunits that form the channel. The functions $\alpha_x = \alpha_x(V_{Hj})$ and $\beta_x = \beta_x(V_{Hj})$ are given in Table 8.1.

As stated in the previous section, the synaptic currents are analogous to the ionic currents, although the corresponding conductances are constructed in a somewhat different way:

$$I^j_{HE} = (E_E - V_{Hj})g^j_{HE}S(V_{H(j-1)}) \,, \tag{8.12}$$

$$I^j_{HI} = (E_I - V_{Hj})g^j_{HI}S(V_{HI}) \,. \tag{8.13}$$

Here, the function $S(V_{pre})$, which depends on the membrane voltage of the presynaptic neuron V_{pre}, represents the fraction of postsynaptic receptor channels that are open in response to the binding of excitatory or inhibitory neurotransmitters, and g_{HE} and g_{HI} are the (constant) maximum conductances of these channels. $S(V_{pre})$ is taken to satisfy the first-order kinetics

$$\frac{dS}{dt} = (1 - S)\alpha_S(V_{pre}) - 0.2275S \,, \tag{8.14}$$

where the function $\alpha_S(V_{pre})$ is given in Table 8.1. The equation for $S(V_{pre})$ has the property of having two different time constants: the fraction of open postsynaptic receptor channels rises with a larger time constant when the presynaptic voltage rises, as during an action potential, and decreases with a smaller time constant when the presynaptic voltage decreases to its resting value.

In addition to the excitatory RA-projecting neurons, our minimal set of HVC units contains an inhibitory neuron. Analogously to the excitatory neurons, the membrane voltage $V_{HI}(t)$ of the inhibitory neuron satisfies an HH equation,

$$C_M\frac{dV_{HI}}{dt} = I^{HI}_L + I^{HI}_{Na} + I^{HI}_K + I^{HI}_{HE} \,. \tag{8.15}$$

Leak, sodium and potassium currents are considered, as before:

$$I^{HI}_L = (E_L - V_{HI})G^{HI}_L \,, \tag{8.16}$$

$$I^{HI}_{Na} = (E_{Na} - V_{HI})G^{HI}_{Na} \,, \tag{8.17}$$

$$I^{HI}_K = (E_K - V_{HI})G^{HI}_K \,, \tag{8.18}$$

Table 8.1. Summary of constants and functions for the HVC–RA model. $v = V - E_{rest}$ is the deviation of the membrane potential from its resting value (in units of mV); the functions α_x and β_x are in units of ms^{-1}; the conductances g, Γ_0 and γ_0 are in units of mS/cm^2; the potentials E_x and V_{pre} are in units of mV; and the current I_{DC} is in units of $\mu\text{A/cm}^2$

$$
\begin{aligned}
\alpha_m &= 0.1\frac{25-v}{\exp\left((25-v)/10\right)-1} \\
\beta_m &= 4\exp\left(\frac{-v}{18}\right) \\
\alpha_h &= 0.07\exp\left(\frac{-v}{20}\right) \\
\beta_h &= \frac{1}{\exp\left((30-v)/10\right)+1} \\
\alpha_n &= 0.01\frac{10-v}{\exp\left((10-v)/10\right)-1} \\
\beta_h &= 0.125\exp\left(\frac{-v}{80}\right) \\
\alpha_S &= 0.15\frac{1}{1+\exp(10-V_{pre})} \\
g_{Na} &= 120 \\
g_K &= 36 \\
g_L &= 0.3 \\
g_{HE}^{j} &= 17.7 \\
g_{HI}^{j} &= 7.5 \\
g_{HE}^{HI} &= 10.5 \\
g_{RI}^{k} &= 75 \\
g_{RE}^{RI} &= 7.5 \\
g_{HE}^{RI} &= 5 \\
\Gamma_0 &= 18.62 \\
\gamma_0 &= 3.5 \\
E_L &= -64 \\
E_{Na} &= 50 \\
E_K &= -95 \\
E_E &= 0 \\
E_I &= -80 \\
I_{DC} &= 1.93
\end{aligned}
$$

where G_L^{HI} is the constant conductance for the leak current. The voltage-dependent sodium and potassium conductances are given by

$$G_{Na}^{HI} = g_{Na} m_{HI}^3 h_{HI} \,, \tag{8.19}$$

$$G_K^{HI} = g_K n_{HI}^4 \,, \tag{8.20}$$

where the dynamics of the activation and inactivation variables m_{HI}, n_{HI} and h_{HI} are given by (8.11). The last term on the right of (8.15) represents the input to the inhibitory unit from all the excitatory HVC units in the model. It is a sum of synaptic currents and is given by

$$I_{HE}^{HI} = (E_E - V_{HI}) g_{HE}^{HI} \sum_{j=1}^{N_{HVC}} S(V_{Hj}) \,. \tag{8.21}$$

The resting potential of the isolated inhibitory HVC neuron is $\sim -65\,\mathrm{mV}$.

Equations (8.5)–(8.21) describe a set of excitatory HVC neurons which are connected in a feed-forward fashion (so that activation of neuron $j = 1$ excites $j = 2$ and so forth to $j = N_{HVC}$, with $S(V_{H0}) = 0$), plus an inhibitory neuron connected to each of them. We use constants guaranteeing that an initial excitation of the first unit will excite the remaining units sequentially. When $g_{HI} \neq 0$, the inhibitory neuron receives excitatory input in a global fashion from the excitatory HVC neurons, and each excitatory HVC neuron receives inhibition as well as sequential excitation. The parameters of the system are adjusted so that the activity of an RA-projecting HVC unit consists of 4 to 7 spikes [Hahnloser et al. 2002].

8.4.2 Simulating the Activity of RA Neurons

A similar procedure can be carried out for the simulation of the activities of the RA neurons. The kth excitatory RA neuron also satisfies an HH equation, with the addition of input synaptic currents from the HVC population and from the inhibitory RA interneuron, as well as excitatory connections from other excitatory neurons within RA. The membrane voltage of the kth RA excitatory neuron, $V_{Rk}(t)$, where $k = 1, 2, \ldots, N_{RA}$, satisfies

$$C_M \frac{dV_{Rk}}{dt} = I_L^k + I_{Na}^k + I_K^k + I_{RI}^k + I_{RE}^k + I_{HE}^k + I_{DC} \,, \tag{8.22}$$

where, again, we have the usual leak, sodium and potassium currents, as well as the synaptic currents I_{RI}^k from the single inhibitory RA neuron and I_{RE}^k from several other RA excitatory neurons. In addition, HVC $\rightarrow$ RA connections are considered: input from several excitatory HVC neurons is represented by the synaptic current I_{HE}^k. With the constant current I_{DC} also present, these excitatory units in RA display, before HVC activity starts, spontaneous spiking at approximately 20 Hz.

The input from the single inhibitory RA neuron, analogously to (8.13), is given by

$$I_{RI}^{k} = (E_I - V_{Rk}) g_{RI}^{k} S(V_{RI}) \,, \tag{8.23}$$

while the input from other excitatory RA neurons is built up as a sum of synaptic currents like those given by (8.12):

$$I_{RE}^{k} = \sum_{k'=1;k'\neq k}^{N_{RA}} \Gamma_{kk'} S(V_{Rk'})(E_E - V_{Rk}) \,. \tag{8.24}$$

The matrix $\Gamma_{kk'}$ determines the strength of the connections among excitatory neurons within RA. Following Spiro and coworkers [Spiro et al. 1999], we consider only matrix elements between closely neighboring units to be different from zero (i.e., units within the same subpopulation of units that controls either tension or respiratory activity). This reflects the local nature of the connectivity of excitatory units within RA. In order to account for this, we take $\Gamma_{kk'} = \Gamma_0 A_{kk'}$, with a constant Γ_0 and $A_{kk'}$ being a matrix with entries of 0 and 1.

Analogously, input from excitatory HVC neurons is taken into account by means of the term in (8.22):

$$I_{HE}^{k} = \sum_{j=1}^{N_{HVC}} \gamma_{kj} S(V_{Hj})(E_E - V_{Rk}) \,. \tag{8.25}$$

The matrix γ_{kj} dictates the strength of the synaptic connections between HVC and RA neurons, and is given by $\gamma_{kj} = \gamma_0 B_{kj}$, with a constant γ_0 and B_{kj} being a matrix with entries of 0 and 1. These connections have been carefully studied, and the results reported in the literature. Optical methods have been applied to describe how these connections evolve and are rearranged as learning takes place [Herrman and Arnold 1991]. More recently, the precise receptors used in the connection from HVC to RA have been described [Stark and Perkel 1999]. This study indicates that both AMPA and NMDA receptors play a role in the connectivity between HVC and RA, although the contribution of NMDA-mediated synaptic currents decreases as maturation takes place.

The membrane voltage V_{RI} of the single inhibitory RA neuron is given by the following HH equation:

$$C_M \frac{dV_{RI}}{dt} = I_L^{RI} + I_{Na}^{RI} + I_K^{RI} + I_{RE}^{RI} + I_{HE}^{RI} \,. \tag{8.26}$$

The usual leak, sodium and potassium currents are considered here. Connections from all excitatory RA neurons and all excitatory HVC neurons are represented by

$$I_{RE}^{RI} = (E_E - V_{RI}) g_{RE}^{RI} \sum_{k=1}^{N_{RA}} S(V_{Rk}) \tag{8.27}$$

and

$$I_{HE}^{RI} = (E_E - V_{RI}) g_{HE}^{RI} \sum_{j=1}^{N_{HVC}} S(V_{Hj}) \,, \tag{8.28}$$

respectively, which are analogous to (8.21).

Equations (8.22)–(8.28) thus describe a set of interconnected RA excitatory neurons which are driven by connections from HVC neurons and are inhibited by the inhibitory RA neuron. The inhibitory RA neuron, in turn, is driven by synaptic input from all the RA excitatory neurons, as well as by projection neurons in HVC.

8.4.3 Qualitative Predictions

In principle, if the purpose of the computational efforts in this field was to reproduce observed behavior, one could argue against their utility. However, models are useful when they allow us to go beyond integrating disparate data. They are useful when they allow us to predict new dynamical regimes, or behavior in regions of parameter space unexplored experimentally. Ideally, they could point to situations worth exploring. The paramount complexity that can in principle be expected in networks of coupled excitable units allows us to predict that the activity in this field has hardly started.

The authors of [Abarbanel et al. 2004a] explored the dynamics displayed by this mathematical model of the simple neural circuit used to represent the motor pathway. In particular, they found that when the syllabic frequency was very high, the average activity of RA (obtained after simulating the dynamics of individual neurons by means of conductance models) showed subharmonics. Abarbanel and coworkers described a period-two solution (i.e., the patterns of neural activity in the nucleus repeat themselves after twice the period of the forcing that activates the neural circuit representing HVC). If the respiratory nuclei and nXIIts nucleus were simply to follow the activity of the different subpopulations of RA that project onto them, we could link this result with the observation that in some species there is an alternation of very similar syllables [Laje and Mindlin 2002]. In fact, this is not necessarily the case. These nuclei are also built from excitable units, and they are interconnected. But these models have predicted complex dynamic responses emerging from interaction between simple instructions and a complex connectivity of nonlinear units.

8.5 Sensorimotor Control of Singing

Between the song box that we have discussed throughout this book and the forebrain nuclei that we described in the previous sections, there is still a set of nuclei that constitute the brain-stem respiratory–vocal system [Wild 2004].

One of these nuclei is the tracheosyringeal motor nucleus (nXIIts), which innervates the syrinx. A second nucleus of this set is the retroambigualis nucleus (RAm), which innervates both nXIIts and the spinal motor neurons which control the expiratory muscles. In birds, both inspiration and expiration are correlated with the activity of specific sets of muscles. The parambigualis nucleus (PAm) has neurons with activity correlated with inspiration [Wild 2004]. How do these nuclei interact? On one hand, the nuclei associated with inspiration and expiration (PAm and RAm) will inhibit each other. On the other hand, there is feedback from the air sacs to these nuclei via the vagus nerve (NX). Intrapulmonary CO_2 receptors are innervated by vagal afferents, as well as sensors of air sac volume [Wild 2004]. The vagal nerve then projects to the tractus solitarius nucleus, which then projects to PAm. We describe this complex structure to point out that the song system should not be thought of as a music box, where a neural script forces a passive peripheral system. On the contrary, the brain-stem nuclei are modulated on one hand by the forebrain nuclei, and on the other hand by peripheral inputs that act in a feedback manner. Experience has shown that forced nonlinear systems can display quite complex behavior. This should prepare us to expect a rich class of phenomena and sound features which are not just the result of a complex musical script written in forebrain nuclei, but are a result of an interaction between the script and the instrument itself.

8.6 Computational Models and Learning

The genetic constraints on song are sufficiently loose in oscine songbirds, hummingbirds and parrots that song acquisition requires an imitation process [Nottebohm 2002b]. However, the processes involved in learning are not completely known. In this area, we are not referring to details: there is a wide range of phenomena which probably play complementary roles in the process of learning, beyond reinforcement or weakening of connections between the neurons of the motor pathway [Nordeen and Nordeen 1997].

Maybe one of the most surprising processes involved in learning is neurogenesis, i.e., the birth of neurons. Biologists usually assume that this occurs in the developing embryo or at a young age. One of the scientific statements that biologists have conveyed to the general public is that neurons in adults cannot be replaced if damaged. For this reason, the observation that Nottebohm and coworkers reported in the 1980s was revolutionary [Nottebohm et al. 1986, Nottebohm 2002a]: they described anatomical changes in the brain which were correlated with developmental changes in song. New neurons were being detected in the brains of adult canaries. But what was really suggestive was that Nottebohm and Alvarez Buylla were able to follow the geometrical paths within the brain followed by these new neurons, and found that they migrated precisely to HVC, one of the nuclei of the motor song system [Alvarez-Buylla and Nottebohm 1988]. A fraction

of the migrating cells reached the target and developed functional connections to other neurons in HVC [Burd and Nottebohm 1985]. To make the picture even more suggestive, many of these new neurons developed projections to RA [Nordeen and Nordeen 1988]. As Catchpole and Slater point out [Catchpole and Slater 1995], there are species in which neurogenesis is not accompanied by changes in the song, as is the case with the zebra finch. Unquestionably, there is an important mechanism to be revealed relating neurogenesis and learning. However, it might not be the only neural process involved in learning.

As we have discussed, there is an indirect connection between the HVC and RA nuclei. It is known as the anterior forebrain pathway, and does not participate in the production of song, but plays a crucial role in learning. The output of this pathway is the lateral magnocellular nucleus of the anterior neostriatum. A series of experiments has suggested that this nucleus responds if some error is detected [Bottjer 2002]. Individual RA neurons receive inputs from both lMAN and HVC, which is consistent with the picture that experience-related lMAN activity facilitates certain HVC–RA synapses, helping to build the neural architecture necessary to produce the adult song.

According to this picture, a sequence of bursts generated at an RA-projecting HVC neuron will induce some activity in RA, and also eventually induce an activity in lMAN that will lead to either the potentiation or the depression of the connection. This signal, however, requires a time for processing through the AFP, which has been estimated as approximately 40 ms [Kimpo et al. 2003]. Abarbanel and coworkers [Abarbanel et al. 2004b] presented a biophysical model for the heterosynaptic facilitation mechanism described above, assuming a sparse activity in the RA-projecting HVC neurons like that reported in the literature [Hahnloser et al. 2002]. Abarbanel and coworkers showed that long term potentiation (LTP) is possible precisely if the delay between the presynaptic activity of the HVC connection and the presynaptic activity arriving from lMAN is less than or equal to 40 ms. Therefore, the estimated processing times of the AFP allow the system to operate in a regime in which either LTP or long term depression (LTD) can be easily achieved through minimal dynamical changes. The basic idea is that a change in calcium concentration in an RA cell can lead to changes in the conductance of the HVC–RA synapses, and that this change in concentration can be induced by the arrival of bursts of action potentials from HVC and lMAN, separated by specific intervals of time (ΔT). The model is computational: it describes the time evolution of the membrane potential of an RA neuron by a conductance model

$$C_M \frac{dV}{dt} = g_L(E_L - V(t)) + I_{HVC}^{NMDA} + I_{HVC}^{AMPA} + I_{lMAN}^{NMDA}\,, \tag{8.29}$$

where the "leak" current drives the neuron to $V = E_L$ in the absence of signaling from HVC or lMAN, C_M is the membrane capacitance per unit area, and the currents are synaptic currents induced by activity in lMAN or

HVC through specific channels known as AMPA and NMDA. According to numerical simulations, a sparse burst in HVC can induce a depolarization of a postsynaptic RA cell. Later activity in an lMAN neuron connected to the same RA cell can induce the entrance of calcium, as long as the cell is still depolarized (since depolarization unblocks the channels used by the calcium to enter the cell). If the concentration of calcium changes the conductance of the HVC–RA synapses [Abarbanel et al. 2002], this process might also be of importance in the learning process. Suggestively, the model predicts that LTP is possible precisely if the delay between the presynaptic activity of the HVC connection and the presynaptic activity arriving from lMAN is less than or equal to 40 ms: precisely the estimated processing time of the AFP [Abarbanel et al. 2004b]. Since for delay times slightly above this value, there is an interval of delay times for which LTD occurs, the system operates in a regime in which either LTP or LTD can be easily achieved through minimal dynamical changes.

This model illustrates the lights and shadows of computational models. It integrates anatomical and physiological information, and suggests a mechanism for learning: changes in the AFP processing time can strengthen or weaken the connections from HVC to RA. It is worth noticing that a computational model of a realistic learning network mimicking these nuclei does not exist. The nature of the coding of lMAN is not known, and only experiments will enlighten this crucial issue.

8.7 Rate Models

The construction of computational models can also be useful for demonstrating theoretically the plausibility of a given mechanism. This was the case in [Doya and Sejnowski 1995], in which a reinforcement model of birdsong learning was conjectured. In this example, the biology inspires the elements used to build the conjectured algorithmic processes.

Another interesting example of this procedure was carried out by Troyer and Doupe [Troyer and Doupe 2000a, Troyer and Doupe 2000b]. These authors proposed an associational model of birdsong learning, guided by well-characterized functional anatomy. The model was designed to test the plausibility of a precise set of hypotheses, among them that the AFP pathway plays an important role in the comparison between the bird's vocalization and a previously memorized template. The issue is not minor, since this beloved hypothesis is challenged by the important delay that presumptive auditory signals would experience as the AFP processes them. Troyer and Doupe used a computational model to show that, if an internal prediction or "efference copy" is used for comparison instead of actual auditory signals, the delay problem can be overcome. Important simplifications are assumed in this model concerning the encoding of motor and sensory information corresponding to a song, but this will be the case for a long time,

since our knowledge of the coding mechanisms involved is far from complete. In particular, Troyer and Doupe used to represent the brain's neural activity a variable that represents the average activity of a large array of neurons. Models at this level of description are known as "rate models" [Troyer and Doupe 2000a, Troyer and Doupe 2000b].

In general, the "activity" of a neuron is measured in terms of the average number of action potentials generated per unit of time, or *firing rate*. One of the simplest continuous-time models used to describe the dynamics of such a variable is the additive network model, which reads

$$\frac{dx_i}{dt} = -x_i + S\left(\rho_i + \sum c_{ij}x_j\right) , \tag{8.30}$$

where x_i denotes the activity of the ith neuron within a network, ρ_i stands for the external inputs, and the coefficients c_{ij} describe the connections between the ith and jth neurons. The function $S(x)$ is a continuous, monotonically increasing function of its real argument which tends to a saturation value for large arguments, and tends rapidly to zero for arguments smaller than zero. For example, we can use the model $S(x) = 1/(1 + e^{-x})$. The dynamics of this equation are simple enough to understand qualitatively: depending on whether the argument of $S(x)$ is large enough or not, the activity of the ith neuron will converge to the saturation value of S or to zero [Hoppensteadt and Izhikevich 1997].

Under certain conditions of the coupling between the neurons within a nucleus, it is possible to attempt to write down a simple average rate model for the activity of a population of neurons. This was the program of Schuster and Wagner [Schuster et al. 1990], who studied a neural circuit of model neurons whose efferent synapses were either excitatory or inhibitory, in a configuration such that the neurons were densely interconnected on a local scale, but only sparsely connected on a larger scale. Under these conditions, Schuster and Wagner showed that it was possible to derive macroscopic mean-field equations for clusters of neurons.

This simplification of the problem would allow us to write, for example, sets of equations governing the behavior of different populations of neurons within some of the nuclei of the motor pathway. For example, we could describe the average activities E_1 and E_2 of the subpopulations of the RA nucleus projecting to the respiratory center and to nXIIts (controlling the syringeal muscles), respectively, and the activity I of the inhibitory interneurons:

$$\tau_1 \frac{dE_1}{dt} = -E_1 + S(\rho_1 + c_{11}E_1 + c_{13}I) , \tag{8.31}$$

$$\tau_2 \frac{dE_2}{dt} = -E_2 + S(\rho_2 + c_{22}E_2 + c_{23}I) , \tag{8.32}$$

$$\tau_3 \frac{dI}{dt} = -I + S(\rho_3 + c_{31}E_1 + c_{32}E_2 + c_{33}I) , \tag{8.33}$$

where the variables ρ_i stand for the input from the high vocal center, and the τ_i are time-scaling parameters.

It is important to stress that the level of the description chosen in devising a model constrains the kind of questions that it can answer. Of course, if a syllable in birdsong is represented by an oscillation in the average activity of HVC, we shall not be able to study those phenomena related to how the sparsely spiking RA-projecting HVC neurons recruit RA neurons. However, it can warn us about global nontrivial phenomena. For example, under periodic driving from HVC (represented by $\rho_i(t) = \rho_i(t + T)$), the system of (8.33) displays subharmonic behavior. This means that under a driving from HVC of period T, the variables E_i and I might repeat themselves after a time nT, where the value of n depends on the frequency and amplitude of the forcing. This warns us that even at the level of telencephalic activity, diverse patterns of activity can be generated by a unique neural substrate operated under different conditions.

Good models should provide us with new questions (and not only good fits to already observed data). With all its limitations, the fact that a rate model shows this nontrivial behavior should motivate the exploration of more sophisticated models (and, ultimately, the exploration of *real* systems through experimental observations).

So far, we have discussed several levels of description of the dynamics displayed by neural circuits. Maybe more importantly, the motivations behind the studies quoted were different. Some of the efforts concentrated on predicting the dynamics that would emerge from an anatomical substrate. Others concentrated on the plausibility of a given mechanism.

Another interesting motivation for modeling is to understand the function of an observed feature. As an example, Fiete et al. [Fiete et al. 2004] explored the role of the sparseness present in premotor neural codes. As we have already discussed, the RA-projecting neurons in HVC display a sparse bursting activity. In zebra finches, which sing motifs that consist of sequences of different syllables, each RA-projecting neuron in HVC spikes briefly (for about 6 ms), once per motif. Fiete and coworkers explored numerically and analytically a network of sparsely spiking units (emulating the dynamics of each unit by a rate model), connected to a set of units emulating the RA nucleus. The motor output of the network was assumed to be some function of the RA activity. With this setup, they explored the efficiency of a learning scheme as a function of the number of bursts in HVC per motif. The learning scheme assumed the existence of a desired motor output, and that at some level, a network error C could be computed. With a backpropagation gradient descent rule, the changes in the synaptic weights between HVC and RA were computed:

$$\Delta W_{ij} = -\epsilon \frac{\partial C}{\partial W_{ij}} \,. \tag{8.34}$$

Within this framework, the time taken to reduce C below a threshold was computed as a function of B, the number of bursts in HVC per motif. The result was that the learning speed decreased dramatically as a function of B; this was found heuristically to be a consequence of the increasing interference of weight updates for different synapses.

8.8 Lights and Shadows of Modeling Brain Activity

The brain is probably the most complex "device" found in nature. Are we ever going to be able to model its dynamics? The brains of songbirds are a tempting field. The number of nuclei involved in the production of song is relatively small, as well as the number of nuclei constituting the "learning" pathway. Our knowledge of the roles played by these nuclei is growing rapidly, as well as our knowledge of some detailed physiological features of the neurons within them.

On the other hand, thousands of nonlinear units coupled through a breathtaking number of connections are the physical substrate of this "device". How complex should our description of an individual unit be? How many units are a reasonable number for an emulation of the activity of a nucleus? Our discussion of the complexity displayed by the physical apparatus, when nonlinear effects are taken into account in models involving a few variables, gives us an intuition about the difficulties of this program. There are no easy answers to these questions. There is a wide consensus that conductance models (such as those proposed by Hodgkin and Huxley) are an appropriate point to start. Dynamicists fight for the hypothesis that simpler models that capture basic dynamical properties of these equations [Izhikevich 2005] can be used instead of complex conductance-based models, and some physicists are testing this hypothesis by replacing real neurons by hardware devices obeying simple dynamical rules [Szucs et al. 2000, Aliaga et al. 2003]. The problem of the proper dimensionality in which to study these issues is not much easier. There is no qualitative theory of extended nonlinear problems. On the other hand, in many cases, we can try to build a consistent story with data from a few neurons.

Quantification in this discipline will not be an easy adventure. Other disciplines, such as physics, have built confidence in general laws that allow us to generate knowledge by theoretically exploring their consequences. Moreover, it is always possible in physics to isolate a phenomenon and study it under the simplest possible conditions. None of these basic pillars is available to the modeler in theoretical neuroscience. However, it is the overwhelming interest of the questions that keeps research in this field alive.

9 Complex Rhythms

There is more to birdsong than just a beautiful melody. So far, we have found "melodies", or sequences of more or less complex sounds. We have focused on the acoustic properties of the individual elements. Now we turn to another aspect of birdsong production: its *rhythm*. Here we use "rhythm" in a very musical sense: we mean "the pattern of musical movement through time" [American Heritage Dictionary 2000], or the timing of sounds within the melody. As we shall see, complex rhythms in birdsong may be found in the song of a single bird or a duetting couple. How are they generated?

9.1 Linear vs. Nonlinear *Forced* Oscillators

We can obtain a precise image of what we mean by *complex* rhythms by simply hitting the table periodically with our left hand (which will behave as a clock in this discussion) while also hitting the table, at a variety of frequencies, with our right hand. The variety of possible timings between impacts of the right hand constitute the various rhythms of the problem. If both hands hit the table the same number of times in a given time interval, we say that they are locked into a "period one" solution (a particular way of achieving "period one" is by always hitting the table with both hands at the same time). Let us now hit the table with our right hand only once per two impacts of the left hand. This is a "period two" solution–the complete pattern does not repeat itself until the left hand has hit the table twice. Now, a more complex excercise. Let us generate a "period three" solution. At this point we have a choice: the right hand can hit the table only once in the time interval that it takes our left hand to hit the table three times, quite analogously to the previous exercise for "period two". But we could also generate a "period three" solution in which the right hand hits the table twice in the time it takes the left hand to hit it three times. These are two different "period three" solutions. In order to distinguish them, one defines the *rotation number* $r = p/q$, where q is the period of the solution, and p the number of recurrences performed by the driven system. In our two examples of "period three" solutions, we would have $r = 1/3$ and $r = 2/3$, respectively. Can we generate these different rhythms with simple coupled physical systems?

We can build a physical system displaying these timings if it is made of two parts: one behaving as a clock, and another one that is forced by the first one. The kind of responses that the driven system might exhibit do not depend on its details, but on its linear/nonlinear nature.

A linear oscillating system responds to a periodic forcing in a remarkably boring way: after a transient, the forced system ends up following the driver. The amplitude of the response will in fact depend on how similar the natural frequency of the system and the forcing frequency are. The phase difference between the oscillations will depend on the parameters, but the driven system will always end up oscillating periodically, with the same period as the driver.

Nonlinear systems, on the other hand, react in a very different way to a periodic forcing. If the forcing frequency is similar to the natural frequency of the system, the forced system will in fact lock itself to the driver in a period one state. But if the difference between these frequencies is large, then the nonlinear system will show a more complex time evolution. The nonlinear system might display periodic behavior but the period of the solution need not, in principle, be equal to the period of the forcing. A typical response of a nonlinear system is to show, for wide regions of the parameter space (in this case, the amplitude and frequency of the forcing), periodic solutions with periods that are multiples of the driving period. This means that the system will repeat its behavior after a time that is an integer multiple of the forcing period.

A neuron, with no musical talent at all, will react to periodic forcing with these kinds of solutions. The typical time it takes a neuron to return to its rest value after an action potential defines a characteristic time. If a neuron is periodically forced by a sequence of pulses with a period comparable to that characteristic time, it will spike with one action potential per period of the forcing. However, if the neuron is forced twice as fast its characteristic time, the neuron will spike only *once* per two periods of the forcing. Two interesting observations: it is not necessary to force the neuron exactly twice as fast as its characteristic time to lock the neuron into a period of two. There is a range of parameter values for which the system locks into a "period two" solution. For the same value of the forcing amplitude, we can change the forcing frequency within a certain range and stay locked in the same regime. The second observation is maybe even more curious. There is a wide range of forcing frequencies in which the neuron locks into a "period one" solution (with $r = p_1/q_1 = 1/1$). There is also a wide range of forcing frequencies in which the neuron displays a "period two" solution (with $r = p_2/q_2 = 1/2$). Now, for a large class of nonlinear systems (excitable ones such as a neuron among them), the largest region in frequency space between these two frequency ranges for which a periodic solution exists is one in which a "period three" solution occurs, characterized by

$$r = \frac{p_1 + p_2}{q_1 + q_2} = \frac{2}{3}\,. \tag{9.1}$$

This is a universal feature of nonlinear oscillators known as the *Farey sum rule*. There is nothing special about neurons. Any excitable system would show such a structure, as well as any nonlinear oscillator. For example, two subpopulations of interconnected excitatory and inhibitory neurons might have excitatory behavior as a whole. By this we mean that the activity of the subpopulation (defined as the average number of spikes in some time window for the subpopulation) can behave as an excitable system [Hoppensteadt and Izhikevich 1997]. For this reason, under a periodic forcing, a whole neural nucleus could display this complex rhythmic structure. In summary, nonlinear systems are capable of generating complex and appealing rhythms. However, these temporal patterns are seriously constrained by a mathematical structure. What about the beautiful rhythms present in birdsong?

9.2 Duets

9.2.1 Hornero Duets

The South American hornero (*Furnarius rufus*) is a suboscine bird, widely known for its nest, which is made of mud and looks like an oven. Beyond their architectural skills, male and female horneros engage in highly structured duets. A sonogram of a typical duet is displayed in Fig. 9.1a, where the continuous traces represent notes. The male starts singing at a note production rate of approximately 6 Hz [Amador 2004], and in a few seconds it increases the note production rate by about 200 percent. The duration of the notes decreases. The female shows a large diversity: it may sing with an increasing, decreasing or nonmonotonically varying note production rate, as can be seen in Fig. 9.2. At the begining of the song (which can last up to 10 seconds), the female is capable of following the male, singing a note each time the male does in a "one-by-one" fashion. However, after a while the female seems to lose synchrony. Nevertheless, the timing has a nontrivial structure, with features characteristic of nonlinear forced oscillators such as the one that we discussed in the previous section.

In [Laje and Mindlin 2003], sonograms of hornero duets were computed from field recordings, and only the fundamental frequencies were displayed, as shown in Fig. 9.1a. The time intervals in which two notes are present are those in which the two duetting birds are vocalizing simultaneously. In order to characterize the duet, we define a *coincidence* as an event in which the maximum of a male note occurs within a time interval in which the female is vocalizing a note. We then define a number that describes the locking between the male and female voices. This number r_{ap} is the quotient of two integers p/q, where q stands for the number of male notes between consecutive male–female coincidences, and p is the number of female notes between consecutive male–female coincidences.

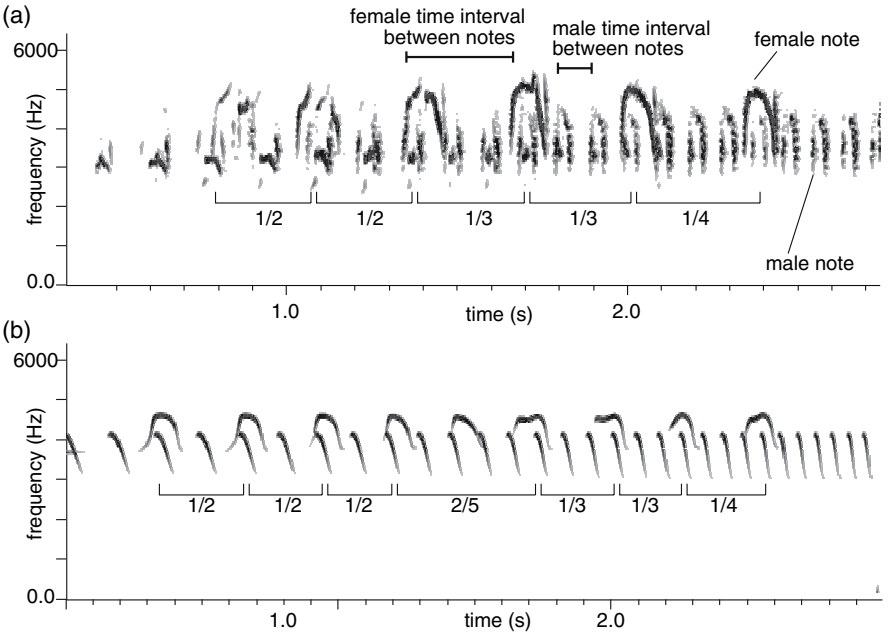

Fig. 9.1. Duet sonograms. (**a**) Sonogram of a typical duet of the suboscine hornero (*Furnarius rufus*). Male and female voices are easily distinguished, owing to the differences in their spectral and temporal features. The male begins the vocalization, monotonically increasing the note production rate as the duet develops. The female, on the other hand, may respond with an increasing, decreasing or nonmonotonically varying note production rate, but their notes lock in a systematic way. The locking sequence of this duet is shown in the sonogram. (**b**) Sonogram of a synthetic duet. Female syringeal and respiratory gestures were simulated by a nonlinear oscillator subjected to a periodic forcing of increasing frequency. Both a periodic forcing and a forced nonlinear oscillator were used to drive physical models of the syrinx to generate male and female notes. Adapted from [Laje and Mindlin 2003]

9.2.2 A Devil's Staircase

In Fig. 9.3a, we plot r_{ap} as a function of the normalized average time interval between male notes T_{male} for $n = 11$ duets (from at least 10 hornero couples), in order to inspect the female response. The average time interval between male notes has been normalized to the average time interval between notes of the corresponding female when locked with $r_{ap} = 1/3$ (every female locked with $r_{ap} = 1/3$ at least once). A clear stair-like structure emerges, with steps at several values of p/q. Notice that some values of p are different from 1. These are located at the regions predicted by the Farey sum rule [González and Piro 1983]. For example, we found a number $r_{ap} = 2/7$; it fell between $r_{ap} = 1/3$ and $r_{ap} = 1/4$. A segment with $r_{ap} = 3/10$ was also found, located between $r_{ap} = 1/3$ and $r_{ap} = 2/7$.

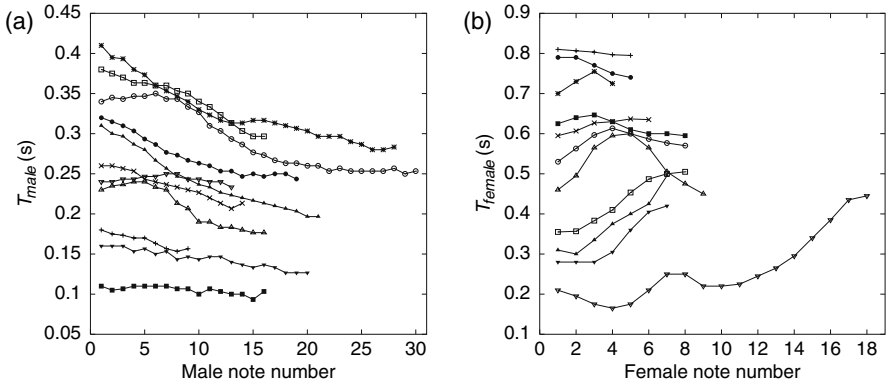

Fig. 9.2. Average time interval between notes as a function of note position in a typical hornero duet. (**a**) Males show a clear tendency to increase the note production rate as the duet develops. (**b**) Females, on the other hand, may increase, decrease or nonmonotonically vary the note production rate as the duet develops. Each trace was shifted vertically for clarity

This structure is very closely related to the one we described at the beginning of the chapter. The organization of the rotation numbers as a function of the forcing period is known as the "devil's staircase", a complex structure where the steps are rational numbers, as displayed in Fig. 9.3b. The steps in the devil's staircase show universal features, regardless of the details of the forced system (which can be either a nonlinear oscillator or an excitable system) and the nature of the periodic forcing (a harmonic function, a sequence of impulses, etc.).

It is possible to interpret the existence of a staircase in the duets as a consequence of the male's notes acting as a unidirectional forcing on the female's motor control through an auditory pathway. Since the notes sung by the female manage to lock properly, both the respiratory and the syringeal gestures would constitute a nonlinear system such that when it was forced it would show a devil's staircase (it would be an excitable system or a relaxation oscillator), suggesting that at some level the corresponding motor neurons are closely connected.

9.2.3 Test Duets

In order to check the plausibility of the proposed mechanism, we can generate synthetic duets from a theoretical, physical model of the syrinx such as the one proposed in Chap. 4, reading

$$\ddot{x} - (p(t) - b)\,\dot{x} + k(t)\,x + cx^2\dot{x} = 0\,, \tag{9.2}$$

where $p(t) - b$ stands for the difference between the bronchial pressure and the linear dissipation, $k(t)$ for the stiffness of the syringeal labia and c stands for

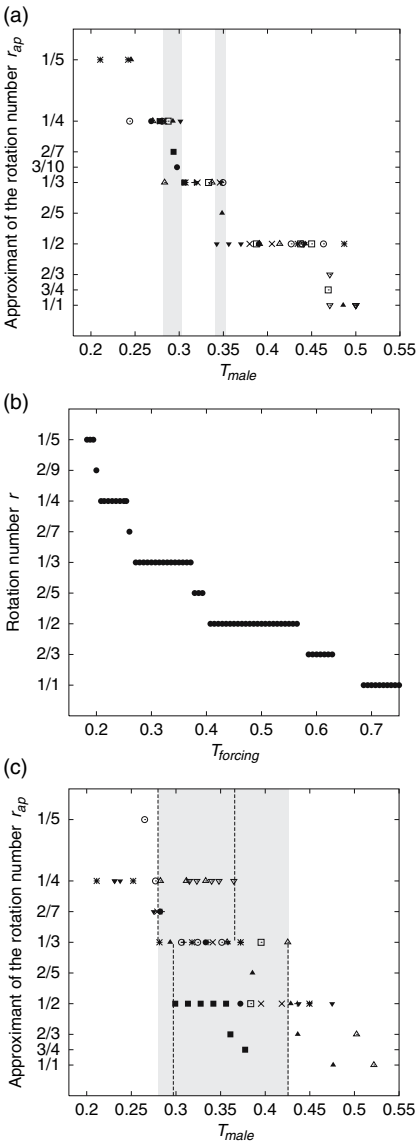

Fig. 9.3. Locking organization. (**a**) Approximant of the rotation number r_{ap} as a function of the normalized average time interval between male notes T_{male}, for $n = 11$ hornero duets. The bottom axis has been normalized for each male to the average time interval between notes of the corresponding female when locked with $r_{ap} = 1/3$. Not every couple displayed the complete locking sequence from 1/1 to 1/5, but all of them displayed 1/3 at least once. (**b**) Rotation number r as a function of forcing period $T_{forcing}$, for a nonlinear oscillator subjected to a periodic forcing. This step-like organization is known as the "devil's staircase". The bottom axis is normalized to the natural period of the driven oscillator. (**c**) Surrogate random duets. The step-like structure is lost, which is reflected in the increased overlapping between steps. This can be quantified by defining the *average overlap* ϕ between consecutive steps (see text): the original duets have $\phi \sim 20\%$, while the surrogate duets have $\phi \sim 70\%$. Overlaps between steps are shown in grey

a nonlinear dissipation constant, all per unit mass of the labia. This equation was first proposed to model labial oscillations in oscines. For suboscines, the tracheal syrinx will show important differences, and actually the membrane tension can be partially correlated with the pressure [Elemans 2004]. At this point the model is just a very crude tool to emulate notes. We can do this by driving (9.2) with a periodic forcing of increasing frequency. For the male, we chose $p(t)$ and $k(t)$ to be harmonic functions of time, with a frequency

increasing from around 5 Hz up to 20 Hz. We call $p(t)$ and $k(t)$ the respiratory and syringeal gestures, respectively.

Analogously, we can generate female notes by driving (9.2) with respiratory and syringeal gestures $p(t)$ and $k(t)$, generated this time by a nonlinear oscillator forced, in turn, by the male gestures. We are not making any assumptions about the biological substrate of these nonlinear oscillators. We do know that the cyclic gestures generating the syllables of the females have to emerge out of some circuitry which involves nonlinear units. We conjecture that this circuitry has to be affected by the auditory signals from the duetting partner. Before attempting to build a biological model for this circuitry, we shall assume that the cyclic gestures are generated by the simplest conceivable nonlinear oscillator, forced by a temporal function of increasing frequency. The paradigmatic model chosen to emulate the cyclic instructions driving the syrinx of the female was

$$\dot{u} = 75(u - u^3 + A\cos(\omega(t)t) + 0.5 - v)\,, \tag{9.3}$$

$$\dot{v} = 6(u + 0.7 - 0.8v)\,, \tag{9.4}$$

with $\omega(t)$ varying from around 5 Hz up to 20 Hz. The cyclic instructions used to drive the syrinx of the female were taken as $p(t) = u - v$ and $k(t) = u + v$, in order to generate realistic-looking syllables. The synthetic duet is displayed in Fig. 9.1b.

OK, but what if the staircase in Fig. 9.3a (the actual duets) was just coincidence? In order to check that it was not an effect of a simple mismatch between frequencies changing independently, surrogate random duets were analyzed. Surrogate duets were assembled by first taking two randomly chosen duets, and then eliminating the male notes from the first duet and the female notes from the second duet. Surrogate duets were then subjected to the same analysis as performed on the original duets. Results are shown in Fig. 9.3c, where the staircase structure is lost. For a quantitative measure of the staircase structure, we defined ϕ as the average overlap between steps, using the steps of 1/2, 1/3 and 1/4 (the only steps for which we know their length approximately know their length). The ratio of the length of the overlap to the average step length was computed between the 1/2 and 1/3 steps, and the same was done between the 1/3 and 1/4 steps. The two ratios were then averaged to give ϕ. In the case of the surrogate duets (Fig. 9.3c), the average overlap ϕ is almost 70%, while in the case of the original duets (Fig. 9.3a) ϕ is less than 20%.

The moral of this tale is that behind the appealing rhythm of the hornero duets, there is more of a complex nonlinear circuitry than "musical talent" (unless one wishes to explore the possibility of musical talent being related to this mechanism!). In any case, just as the acoustic features of the notes (as shown in the spectra) were in some cases affected by nonlinear effects (on the timescale of thousands of Hz), nonlinearities can also play an important role on the timescale of the syllables.

9.3 Nonlinear Dynamics

9.3.1 A Toy Nonlinear Oscillator

The observations described in the previous section motivate us to ask two questions: Why should these temporal patterns be present in forced nonlinear oscillators? If so, what is the oscillator in this system? We begin this section by addressing the first question. We shall do this by solving an example, and illustrating how these temporal patterns emerge. Because biological oscillators are usually stable nonlinear oscillators [Glass 2001], many characteristics of the interactions between a periodic input and an ongoing rhythm can be understood in terms of simple toy models, without paying much attention to the details of the system [Glass and Mackey 1988, Keener and Sneyd 1998]. For this reason, we can choose to explore the origins of our strange temporal patterns in terms of a simple system, learning about the tools and processes that take place in order to gain intuition about the particular system we might be interested in.

The toy model that we shall use is known as the Poincaré oscillator, named after the French mathematician Henri Poincaré, who used it precisely to illustrate general mechanisms of nonlinear systems. This abstract oscillator is described in terms of a radial variable ρ and an angular variable ϕ, whose dynamics are governed by the following system of equations:

$$\begin{aligned} \frac{d\rho}{dt} &= \lambda\rho(1-\rho)\,, \\ \frac{d\phi}{dt} &= \omega \,(\text{mod } 2\pi)\,, \end{aligned} \tag{9.5}$$

where "mod 2π" denotes a normalization of the phase dynamics: when $\phi = 2\pi$, the oscillator has completed a turn in the angular coordinate, and starts again from $\phi = 0$. The dynamics of this system of equations are easy to understand: the phase of the oscillator increases monotonically at a constant rate (of value ω). The behavior of the radial part of the equation is also simple: for initial conditions close to the origin, the radial component will grow with time, as long as $\lambda > 0$. On the other hand, for initial conditions far away from the origin (i.e., with a radial part much larger than 1), the system will evolve so that the radial coordinate decreases (at a rate given approximately by $\lambda\rho^2$). The "equilibrium" is reached for the radial part when $\rho = 1$. In this case, the dynamics are reduced to a monotonic increase of the phase, and after a given amount of time ($t = 2\pi/\omega$), the solution repeats itself. This solution is known as a *limit cycle*. The term "limit" in this case refers to the asymptotic nature of the periodic solution: an initial condition not in the cycle (i.e., either outside or inside) will evolve towards it. Notice that the larger the value of λ, the faster the time evolution of the system will be towards the limit cycle, for initial conditions not in the limit cycle. For our description of forced nonlinear oscillators we shall assume that λ is large, so convergence is fast.

9.3.2 Periodic Forcing

Let us now explore the response of the system to a periodic forcing. Again, we shall use the simplest possible forcing term: a periodic "kick" to the system by impulses applied to one coordinate. It would be tempting to try a sinusoidal forcing (which would of couse also be periodic), but impulses have an advantage: the forcing is simply a displacement of the state of the system each time a kick occurs. Between kicks, the system evolves by the rules (9.5). These rules, on the other hand, give rise to simple dynamics: a rapid decay to $\rho = 1$ (assumed to be almost instantaneous if λ is large), and a free evolution of the phase at a constant rate. In summary, the dynamics of the kicked system are as follows: the system is kicked (by the way, this may cause an important change in the phase of the oscillator), it rapidly evolves back to the limit cycle and freely evolves along it, until it is kicked again. The process then repeats itself while the forcing is "on". Our challenge is to understand how a structure such as that displayed in Fig. 9.3b can emerge out of this procedure.

In Fig. 9.4, we display the processes described in the previous paragraph. The circle represents the unforced limit cycle. The system is assumed to be at the point denoted by 1 in the nth step of our iteration procedure, its state being described by the phase $\phi_n = \phi$. When the system is kicked, by means of a horizontal displacement of value A towards the point denoted by 2, the phase of the system is changed to ϕ_a. The system then collapses rapidly towards the limit cycle (the state denoted by 3 in the figure), finally to evolve freely until its phase has increased by an additional amount α (equal to ωT, where T is the period of the forcing), at point 4. Then, a new kick is applied to the oscillator. Since we are working in the limit of $\lambda \gg 1$ (which means $\rho \sim 1$ almost instantaneously after a kick), the state of our forced oscillator can be described in terms of the dynamics of the phase ϕ_n only, which can then be written as

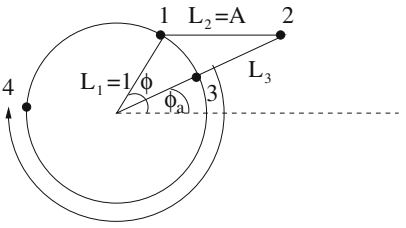

Fig. 9.4. Derivation of the phase map for a Poincaré oscillator, periodically kicked along the horizontal axis. The evolution can be described as the iteration of the following steps: (1) the system is kicked, (2) it rapidly collapses back to the limit cycle, (3) the system evolves with a monotonic increase of the phase and (4) it is kicked again after a time T, where T is the period of the forcing

$$\phi_{n+1} = \alpha + \phi_a(\phi_n, A)\,, \tag{9.6}$$

where ϕ_a is the change in phase that occurs when the system is kicked. The phase ϕ_a depends on the phase ϕ_n of the system at the moment it is kicked and on the size of the kick A. From Fig. 9.4, it is simple to derive an expression for this new phase. Notice that when the system is displaced from point 1 to point 2, we can define a triangle with vertices at the origin and at points 1 and 2. The sides are of lengths $L_1 = \rho = 1$ (the value of the radial coordinate), $L_2 = A$ and $L_3 = \sqrt{A^2 + 1 - 2A\cos(\pi - \phi_n)}$ (by the cosine theorem). Projecting the three sides of our triangle onto the x axis, we can write

$$L_3 \cos(\phi_a(\phi_n, A)) = \cos(\phi_n) + A\,, \tag{9.7}$$

which together with (9.6), gives us an expression for the phase of the oscillator in the nth step as a function of the phase in the previous step. This reads

$$\phi_{n+1} = P(\phi_n) \equiv \alpha + \cos^{-1}\left(\frac{A + \cos(\phi_n)}{\sqrt{A^2 + 1 - 2A\cos(\pi - \phi_n)}}\right), \tag{9.8}$$

which is known as the phase map of the Poincaré oscillator.

9.3.3 Stable Periodic Solutions

Equation (9.8) was easy to find, but it does not look easy to work with. Nevertheless, it can help us in our process of gaining an understanding of how complex patterns are generated by such systems. First of all, we have to learn how to interpret the solutions of these maps in terms of the solutions of the original problem: a forced oscillator. The simplest temporal pattern that could emerge out of the forcing of a nonlinear system is a "one-to-one" locked state: the system repeats its behavior after a time equal to the time between forcing kicks. How would this pattern be represented in the map formalism? Since this formalism only inspects what the system is doing at discrete times, an underlying periodic solution would be captured by a map formalism as a fixed point of the discrete dynamics. In other words, a period-one solution (measured in units of the forcing period) corresponds to a situation in which the phase of the system at a given kick is the same as the phase at the kick before:

$$\phi_{n+1} = \phi_n\,, \tag{9.9}$$

where ϕ_{n+1} is given by (9.8). Notice that this is a transcendental equation and has no explicit solution, but it can be solved approximately either by numerical methods or by graphical means. We should be warned, though, that the existence of a solution does not mean by itself that we are going to observe it (either in an experiment or in a numerical simulation). A solution can exist but be unstable. If we are dealing with a numerical simulation, the slightest difference between an initial condition and the point closest to it

that is part of an unstable solution can grow with time, taking us away from the solution. Therefore, the claim that, for a wide region of the parameter space (T, A), a given model shows solutions locked in a particular way refers to "stable" solutions, i.e., those toward which the system will evolve for all initial conditions within a wide region of the phase space. This requires a condition additional to the one expressed in (9.9), namely one that guarantees stability. The condition is easy to understand: if we expand the map P by calculating its first derivative DP,

$$\begin{aligned} \phi_{n+1} + \delta &= P(\phi_n + \epsilon) \\ &\sim P(\phi_n) + DP(\phi_n)\epsilon\,, \end{aligned} \tag{9.10}$$

the condition that δ is smaller than ϵ is guaranteed if the value of DP is of modulus smaller than one. Now we are able to find the region of parameter space for which stable solutions of period one can exist: we have to show that the cuves $\phi_{n+1} = \phi_n$ cross for some value of ϕ, and that the slope of the curve $\phi_{n+1} = P(\phi_n)$ at the point of intersection is smaller than one.

9.3.4 Locking Organization

What about other temporal patterns? After all, we became involved in this tour of nonlinear phenomena in order to explain "complex" rhythmic patterns. The procedure to find the regions where other solutions exist is very simple: we define a new map which is the second iterate of the one we have used so far. The set of fixed points of the second-iterate map, P^2,

$$\begin{aligned} \phi_{n+2} &= P^2(\phi_n) \\ &= P[P(\phi_n)]\,, \end{aligned} \tag{9.11}$$

is the solutions to the following equation:

$$\phi_{n+2} = \phi_n\,. \tag{9.12}$$

This will contain, on the one hand, the period-one fixed points that we have already met (a solution that repeats itself after one period of the forcing will keep on repeating itself after n times as well). But on the other hand, this set can also contain solutions that repeat themselves after *twice* the period of the forcing, without repeating themselves in one period. These are "period-two" solutions. Working with the map (9.11), we can again impose the conditions of existence and stability of fixed points that we discussed before, and find the region of parameter space where these conditions are met.

The result of performing this study for solutions of different periods is displayed in Fig. 9.5, where the "tongues" enclose the set of parameter values (T, A) for which solutions of different period occur. Moreover, we can advance in our description of the system by defining the *rotation number* r, a ratio of

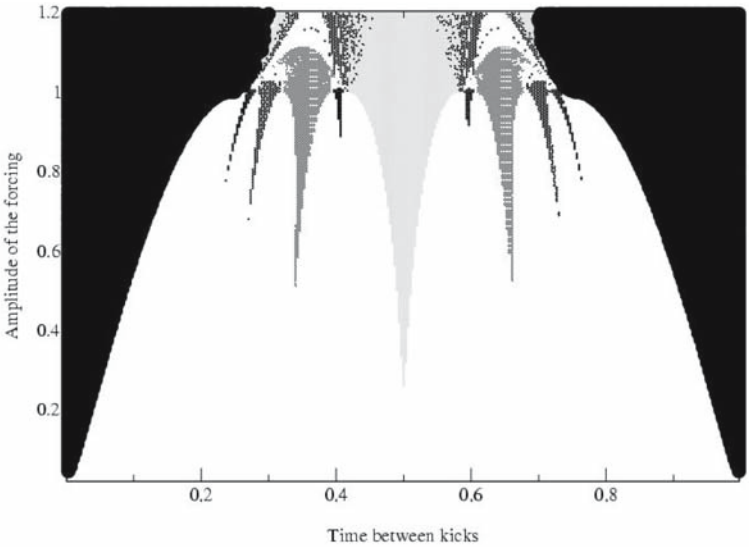

Fig. 9.5. Locking organization for the Poincaré oscillator, periodically kicked along the horizonal axis. The "Arnold tongues" are the regions of the parameter space (T, A) where stable solutions of different rotation numbers can be found, identified here by different gray levels. T = forcing period (in units of the period of the oscillator $T_0 = 2\pi/\omega$); A = forcing amplitude

integers $r = p/q$, where q represents the period of the stable solution (in units of the forcing period), and p the number of complete cycles that the forced system performs before repeating itself. In this way, the period-one solutions of (9.9) can be labeled by $r = 1/1$ (the white region in Fig. 9.5), while the period-two solutions of (9.12) can be labeled by $r = 1/2$ (the first gray level). For a given value of A, we could plot the rotation numbers obtained in successive numerical simulations performed for different forcing periods. The result would be a more or less complex stair, like the one described in Sect. 9.2. But what is the dynamical mechanism behind the existence of these "steps"?

Figure 9.6 allows us to gain some intuition. In this figure, we show the (graphical) solution to the fixed points of the second-iterate map at parameter values corresponding to the begining and the end of the period-two tongue, for a constant amplitude $A = 0.8$. In other words, the intersections between the straight line and the curve represent those values of ϕ fulfilling (9.12). Notice that it is possible that there may be no intersections at all; the existence of intersections is governed by the values of the parameters T and A, which set the position of the curve in the plane. As the forcing frequency is increased, the main change in the curve that can be described is a global shift downwards. However, the curve preserves its qualitative shape, and this is why we have found no intersections for $T < 0.47$, two intersections

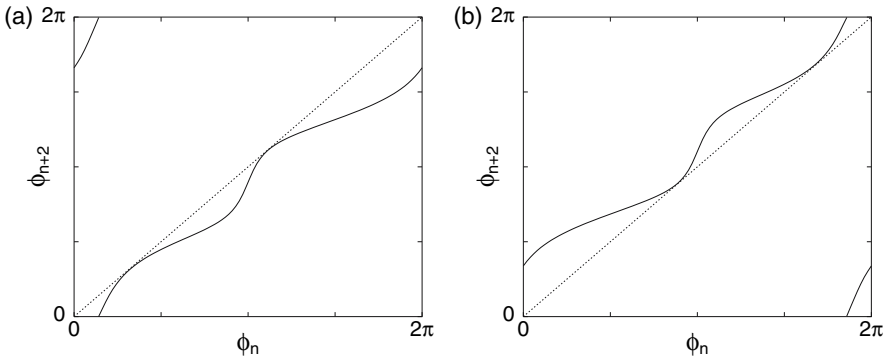

Fig. 9.6. Graphical solution to the fixed points of the second-iterate map (9.12): the *right-hand* side of the equation is represented by the *straight line*, and the *left-hand* side is represented by the *curve*. The parameter values are $A = 0.8$ and (**a**) $T = 0.47$, (**b**) $T = 0.53$ (in units of the period of the oscillator $2\pi/\omega$). As the forcing frequency is decreased, the main change that is observed in the curve is a global shift upwards. These values of forcing period define the borders of the period-two region; for T between 0.47 and 0.53, there are four intersections (solutions to (9.12)), but only two of them represent stable solutions

at $T \sim 0.47$ (Fig. 9.6a), four intersections for $0.47 < T < 0.53$, again two intersections at $T \sim 0.53$ (Fig. 9.6b) and, finally, no intersections for $T > 0.53$. Only two of the four intersections found for $0.47 < T < 0.53$ represent stable solutions. This behavior tells us that the values $T \sim 0.47$ and $T \sim 0.53$ are the borders of the period-two tongue for $A = 0.8$.

It is left for dynamicists to explain what is the minimal set of "geometric" conditions that a one-dimensional map has to satisfy in order to show this organization of temporal patterns. At the rough qualitative level that we are interested in, nonlinear relaxation oscillators and excitable systems do fall within the class.

What is the relationship between this abstract oscillator and the hornero that we met in Sect. 9.2? The hornero is a suboscine. The brain structure in the suboscines is less well understood than that of the oscine songbirds. For this reason, it is a complex task to try to map the abstract oscillator that we are using in this section to a particular structure in the brain. However, the generation of repetitive syllables requires an oscillatory rhythm in order to generate expirations, and therefore some population of neurons (which will eventually feel somatosensory feedback) should show oscillatory activity. The temporal structures described in the section on hornero duets could then be produced if this neural substrate behaved as a nonlinear oscillator and could be forced by auditory signals.

9.4 Respiration

The duets between horneros are a spectacular display of complex rhythms. However, many birds show a wide variety of intersyllabic times. How could these be generated? This section contains a speculative model of their generation, inspired by anatomical descriptions of neural nuclei.

9.4.1 Periodic Stimulation for Respiratory Patterns

In the case of the oscine songbirds, much is known about the structure of the nuclei involved in the generation of song. However, much of the description of the avian nervous system focuses on the forebrain nuclei and pathways [Wild 2004]. The extratelencephalic projections are described as if they were simple followers of the activity in the forebrain nuclei. In fact, this is unlikely to be true, since brain-stem nuclei are receptors of the sensory feedback which informs the song system of the requirements for air. Brain-stem nuclei receive both information from the telencephalic nuclei and sensorimotor feedback, and are therefore a place where very interesting dynamics could be expected. Horneros are suboscines. These birds are believed (after work with the eastern phoebe [Kroodsma and Konishi 1991]) to develop normal song without auditory feedback. They also seem to lack the telencephalic nuclei widely studied in songbirds. They are expected to have a nucleus DM, projecting to the nucleus XIIts, as well as to respiratory nuclei. The respiratory pathways are expected to show similar structures in oscines and suboscines.

Wild and others [Wild 2004, Sturdy et al. 2003] have described the pathways involved in the control of respiration during quiet respiration and singing. The respiratory rhythm is generated or conveyed by the rostral nucleus of the ventrolateral medulla (RVL), which innervates the retroambigualis (RAm) nucleus. This nucleus projects to expiratory motor neurons. On the other hand, the nucleus PAm projects to inspiratory motor neurons, and receives feedback from sensors which update the nervous system about the dynamics of the air sacs (see Fig. 9.7a). The details of the connections are not yet known, but on the basis of this partial list of observations we can build a computational model in order to explore the possible solutions that this respiratory system can display. Interestingly enough, the effects emerging out of this architecture can be present in both oscines and suboscines.

9.4.2 A Model

Recently, Trevisan and coworkers [Trevisan et al. 2005] proposed a simple model to translate these anatomical observations into a computational model. In it, they described the air sac dynamics in terms of a geometric variable measuring the variation x of the air sac volume at atmospheric pressure. The sacs are modeled as damped masses, driven by expiratory and inspiratory muscles (a departure from what happens with mammals, where the

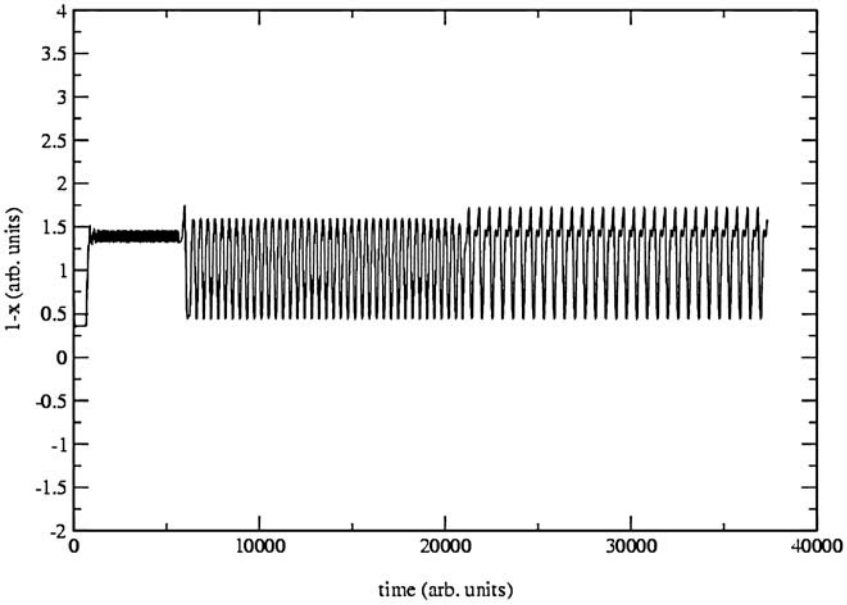

Fig. 9.7. Diverse air sac pressure patterns that can be generated by simple RA activity forcing brainstem nuclei at different frequencies

only driving is inspiratory). If the activities of these muscles are taken to be proportional to the activities in the neural nuclei RAm and PAm (called i_2 and i_1), one can write

$$m\frac{d^2x}{dt^2} + kx + b\frac{dx}{dt} = \alpha_1 i_1 - \alpha_2 i_2 \tag{9.13}$$

for x, where the term kx accounts for elastic recoil of the sac and $b\,dx/dt$ for the dissipation. The activities i_1 and i_2 represent global averages of the activity within the nuclei PAm and RAm. Therefore, their dynamics can be described in terms of average-rate models [Hoppensteadt and Izhikevich 1997]:

$$\begin{aligned} \tau\frac{di_1}{dt} &= -i_1 + S(E_1 - i_2 - f(x))\,, \\ \tau\frac{di_2}{dt} &= -i_2 + S(E_2 - i_1 + A\cos(\omega t))\,. \end{aligned} \tag{9.14}$$

In this model, $S(x) = 1/(1 + e^{-x})$ is a sigmoidal function, typically used to model saturation in neural rate models. The rationale behind the use of this function as a driver of the activities in the nuclei is simple: the greater the amount of excitation that converges on a neuron, the more active it is. However, since beyond a certain excitation the spiking rate of a neuron does not increase further, the mean activity of a neuron has to converge to a saturating

function of the variables in the problem. Assuming that within these nuclei the neurons are densely interconnected on a local scale, we can describe the activity of the nuclei in terms of averaged quantities i_1, i_2 satisfying (9.14) [Schuster et al. 1990].

The arguments of the sigmoidal functions represent the following observation: inspiration and excitation, if coupled, should inhibit each other. RAm receives input from RVL (which we model with an average periodic activity, induced by RA neurons in the case of oscines, and maybe by neurons sensitive to the male song in the brain of female horneros). The function $f(x)$ acts as an inhibitor of inspiration, and mimics the action of stretch receptors or CO_2 receptors in the respiratory system [Keener and Sneyd 1998]. The details of these mechanisms are yet to be discovered, but evidence of somatosensory modulation of ongoing song patterns has been recently presented [Suthers et al. 2002]

As we have discussed, different forcing frequencies can lead to different respiratory patterns. We assume that $1-x$ can represent the air sac pressure (negative x represents small air sacs, and therefore a pressure higher than atmospheric pressure). In Fig. 9.7, we show numerically generated pressure patterns. The three segments correspond to the same neural substrate, but forced with different frequencies (by RA, if we are dealing with oscines). This is a possible mechanism for the generation of diverse temporal sequences. For the moment, this mechanism is no more than a speculative, qualitative one, compatible with the anatomy described so far. However, it highlights the amount of possible dynamics that can emerge from the interaction between the body and the nervous system.

The complexity of respiratory patterns that emerges out of the interaction between neural instructions and the body is a nice example of a rich physiological rhythm. In the literature, several examples have been reported of this kind of dynamical behavior [Glass 2001], and in many cases, explanations have been sought in the richness of solutions of nonlinear systems. What remains rather poorly understood is whether complex dynamics are an essential feature, or a consequence of the processes of the interaction with environment. A mechanism such as the one described in this section highlights the enormous richness of possibilities that a nonlinear substrate provides.

9.5 Body and Brain

Much of the study of the behavior that enhances the survival and reproduction of animals is carried out in terms of its neural control. The emergence of behavior, however, involves a strong interaction betwen the nervous system, the morphology and the environment [Chiel and Beer 1997]. From this point of view, the biomechanics of the peripheral system is a source of opportunities as well as a source of constraints.

Throughout these pages, we have analyzed birdsong in this light. Neural instructions drive a highly nonlinear physical system, the syrinx, capable of generating sounds ranging from simple tones to very complex sounds. Interestingly enough, the complex sounds could be the result of simple instructions, owing to the richness of responses that a nonlinear system can display under forcing. We have described some of the neural processes involved in the song motor pathway. This is a field which is growing rapidly. However, much work needs to be done. Some of the muscle instructions in charge of controlling the vocal organ also emerge from a complex interaction between the body and the neural circuits. Birdsong is becoming a beautiful field for exploring some of these problems, providing biology, physics and dynamics with exciting challenges.

References

[Abarbanel et al. 2002] Abarbanel H. D. I., Huerta R. and Rabinovich M. I. Dynamical model of long-term synaptic plasticity. Proc. Natl. Acad. Sci. USA **99**, 10132–10137 (2002).

[Abarbanel et al. 2004a] Abarbanel H. D. I. A., Gibb L., Mindlin G. B. and Talathi S. Mapping neural architectures onto acoustic features of birdsong. J. Neurophysiol. **92**, 000–000 (2004)

[Abarbanel et al. 2004b] Abarbanel H. D. I., Gibb L., Mindlin G. B., Rabinovich M. I. and Talathi S. Spike timing and synaptic plasticity in the premotor pathway of birdsong. Biol. Cybern. **3**, 1–9 (2004).

[Aliaga et al. 2003] Aliaga J., Busca N., Minces V., Mindlin G. B., Pando B., Salles A. and Szczupak L. Electronic neuron within a ganglion of a leech (*Hirudo medicinalis*) Phys. Rev. E. **67**, 061915 (2003).

[Allan and Suthers 1994] Allan S. E. and Suthers R. A. Lateralization and motor stereotypy of song production in the brown headed cowbird. J. Neurobiol. **25**, 1154–1166 (1994).

[Alvarez-Buylla and Nottebohm 1988] Alvarez-Buylla A. and Nottebohm F. Migration of young neurons in adult avian brain. Nature **335**, 353–354 (1988).

[Amador 2004] Amador A. Ph.D. thesis, Physics Department, UBA, in preparation.

[American Heritage Dictionary 2000] *American Heritage Dictionary of the English Language*, Houghton Mifflin, Boston 4th Edition (2000).

[Arnold et al. 1999] Arnold V. I., Afrajmovich V. S., Ilyashenko Yu. S. and Shilnikov L. P. *Bifurcation Theory and Catastrophe Theory*. Springer, Berlin, Heidelberg (1999).

[Aubier et al. 2000] Aubier T., Jouventin P. and Hildebrand C. Penguins use the two voice system to recognize each other. Proc. R. Soc. London (Biol.) **267**, 1081–1087 (2000).

[Ballintijn and ten Cate 1998] M. R. Ballintijn and C. ten Cate. Sound production in the collared dove: a test of the whistle hypothesis. J. Exp. Biol. **201**, 1637–1649.

[Beckers et al. 2003a] Beckers G. J. L., Suthers R. and ten Cate C. Pure tone birdsong by resonance filtering of harmonic overtones. Proc. Natl. Acad. Sci. USA **100**, 7372–7376 (2003).

[Beckers et al. 2003b] Beckers G. J. L., Suthers R. and ten Cate C. Mechanisms of frequency and amplitude modulation in the ring dove song. J. Exp. Biol. **206**(11), 1833–1843 (2003).

[Bottjer 2002] Bottjer S. Neural strategies for learning during sensitive periods of development. J. Comp. Physiol. A **188**, 917–928 (2002).

[Bottjer et al. 1984] Bottjer S., Miesner E. A., Arnold A. P. Forebrain lesions disrupt development but not maintenance of song in passerine birds. Science **224**, 901–903 (1984).

[Brainard and Doupe 2000] Brainard M. and Doupe A. What birdsong teaches us about learning? Nature **417**, 351–358 (2002).

[Burd and Nottebohm 1985] Burd G. D. and Nottebohm F. Ultrastructural characterization of synaptic terminals formed on newly generated neurons in a song control nucleus of the adult canary forebrain. J. Comp. Neurol. **240**, 143–152 (1985).

[Calder 1970] Calder W. A. Respiration during song in the canary (*Serinus canaria*). Comp. Biochem. Physiol. **32**, 251–258 (1970).

[Casey and Gaunt 1985] Casey R. M. and Gaunt A. S. Theoretical models of the avian syrinx. J. Theor. Biol. **116**, 45–64 (1985).

[Catchpole and Slater 1995] Catchpole C. K. and Slater P. J. B. *Birdsong: Biological Themes and Variations*. Cambridge University Press, Cambridge (1995).

[Chi and Margoliash 2001] Chi Z. and Margoliash D. Temporal precision and temporal drift in brain and behavior of zebra finch song. Neuron **32**, 899–910 (2001).

[Chiel and Beer 1997] Chiel H. and Beer R. D. The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. Trends Neurosci. **20**, 553–557 (1997).

[Doupe and Kuhl 1999] Doupe A. J. and Kuhl P. K. Birdsong and human language: common themes and variations. Annu. Rev. Neurosci. **22**, 567–631 (1999)

[Doya and Sejnowski 1995] Doya K. and Sejnowski T. J. A novel reinforcement model for birdsong vocalization learning. In *Advances in Neural Information Processing Systems*, ed. by Tesaura G., Touretzky, D. S. and Leen, T. K., vol 7. MIT Press, Cambridge, MA (1995), pp. 101–108.

[Elemans 2004] Elemans C. *How Do Birds Sing?*, Ph.D. thesis Leide University, The Netherlands (2004).

[Elemans et al. 2003] Elemans C. P. H., Larsen O. N., Hoffmann M. R. and Leeuwen J. L. Quantitative modeling of the biomechanics of the avian syrinx. Animal Biol. **53**, 183–193 (2003).

[Elemans et al. 2004] Elemans C. P. H., Spierts I. L. Y., Muller U. K., Van Leeuwen J. L. and Goller F. Superfast muscles control dove's trill. Nature **431**, 146 (2004).

[Fee 2002] Fee M. S. Measurement of the linear and nonlinear mechanical properties of the oscine syrinx: implication for function. J. Comp. Physiol. A **188**, 829–839 (2002).

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

[Fee et al. 2004] Fee M. S., Kozhevnikov A. A. and Hahnloser R. H. R. Neural mechanisms of vocal sequence generation in the songbird. Ann. N. Y. Acad. Sci. **1016**, 153–170 (2004).

[Feingold et al. 1988] Feingold M., Gonzalez D., Piro O. and Viturro H. Phase locking, period doubling and chaotic phenomena in externally driven excitable systems. Phys. Rev. A **37**, 4060–4063 (1988).

[Feynman et al. 1970] Feynman R., Leighton R. and Sands M. *The Feynman Lectures Notes on Physics*. Addison Wesley Reading, MA (1970).

[Fiete et al. 2004] Fiete I. R., Hanslosser R. H. R., Fee M. S. and Seung H. S. Temporal sparseness of the premotor drive is important for rapid learning in a neural network model of birdsong. J. Neurophysiol. **92**, 2274–2282 (2004).
[Fletcher 1988] Fletcher N. H. Birdsong: a quantitative acoustic model. J. Theor. Biol. **135**, 455–481 (1988).
[Fletcher 2000] Fletcher N. H. A class of chaotic calls. J. Acoust. Soc. Am. **108**, 821–826 (2000).
[Fletcher and Tarnopolsky 1999] Fletcher N. H. and Tarnopolsky A. Acoustics of the avian vocal tract. J. Acoust. Soc. Am. **105**, 35–49 (1999).
[Fletcher et al. 2004] Fletcher N. H., Riede T., Beckers G. J. L. and Suthers R. Vocal tract filtering of the 'coo' of doves. Preprint (2005).
[Gardner et al. 2001] Gardner T., Cecchi G., Magnasco M., Laje R. and Mindlin G. B. Simple motor gestures for birdsongs. Phys. Rev. Lett. **87**, Art. 2008101, 1–4 (2001).
[Gaunt 1983] Gaunt A. S. An hypothesis concerning the relationship of syringeal structure to vocal abilities. The Auk **100**, 853–862 (1983).
[Gaunt et al. 1982] Gaunt A. S., Gaunt S. L. L. and Casey R. Syringeal mechanisms reassessed: evidence from streptopelia. The Auk **99**, 474–494 (1982).
[Glass 2001] Glass L. Synchronization and rhythmic processes in physiology. Nature **410**, 277–284 (2001).
[Glass and Mackey 1988] Glass L. and Mackey M. C. *From Clocks to Chaos: The Rhythms of Life*. Princeton University Press, Princeton (1988).
[Goller 1998] Goller F. Vocal gymnastics and the bird brain. Nature **395**, 11–12 (1998).
[Goller and Larsen 1997a] Goller F. and Larsen O. N. A new mechanism of sound generation in songbirds. Proc. Natl. Acad. Sci. USA **94**, 14787–14791 (1997).
[Goller and Larsen 1997b] Goller F. and Larsen O. N. In situ biomechanics of the syrinx and sound generation in pigeons. J. Exp. Biol. **200**, 2165–2176 (1997).
[Goller and Larsen 2002] Goller F. and Larsen O. N. New perspectives on mechanisms of sound generation in songbirds. J. Comp. Physiol. A **188**, 841 (2002).
[Goller and Suthers 1995] Goller F. and Suthers R. A. Implications for lateralization of birdsong from unilateral gating of bilateral motor patterns. Nature **373**, 63–66 (1995).
[Goller and Suthers 1996a] Goller F. and Suthers R. A. Role of syringeal muscles in gating airflow and sound production in singing brown thrashers. J. Neurophysiol. **75**, 867–876 (1996).
[Goller and Suthers 1996b] Goller F. and Suthers R. A. Role of syringeal muscles in controlling the phonology of bird song. J. Neurophysiol. **76**, 287–300 (1996).
[Goller and Suthers 1999] Goller F. and Suthers R. A. (1999). Bilaterally symmetrical respiratory activity during lateralized birdsong. J. Neurobiol. **41**, 513–523 (1999).
[González and Piro 1983] González D. L. and Piro O. Chaos in a nonlinear driven oscillator with exact solution. Phys. Rev. Lett. **50**, 870–872 (1983).
[Greenwalt 1968] Greenwalt C. H. *Birdsong: Acoustic and Physiology*. Smithsonian Institute Press, Washington, DC (1968).
[Hahnloser et al. 2002] Hahnloser R. H. R., Kozhevnikov A. A. and Fee M. S. An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature **429**, 65–70 (2002).
[Hartley 1990] Hartley R. S. Expiratory muscle activity during song production in the canary. Respir. Physiol. **81**, 177–188 (1990).

[Hartley and Suthers 1989] Hartley R. S. and Suthers R. A. Airflow and pressure during canary song: evidence for mini-breaths. J. Comp. Physiol. **165**, 15–26 (1989).

[Herrman and Arnold 1991] Herrmann K. and Arnold A. P. The development of afferent projections to the robust archistriatal nucleus in male zebra finches: a quantitative electron microscopic study. J. Neurosci. **11**, 2063–2074 (1991).

[Herzel et al. 1995] Herzel H., Berry D. A., Titze I. R. and Steinecke I. Nonlinear dynamics of the voice: signal analysis and biomechanical modeling. Chaos **5**, 30–34 (1995).

[Herzel et al. 1996] Herzel H. (1996). Possible mechanisms for voice instabilities, in *Vocal Fold Physiology*, ed. by Davis P. J. and Fletcher N. H. Singular Publishing Group, San Diego (1996), pp. 63–75.

[Hildebrand 1995] Hildebrand M. *Analysis of Vertebrate Structure*. Wiley, New York (1995).

[Hodgkin and Huxley 1952] Hodgkin A. L. and Huxley A. F. The components of membrane conductance in the giant axon of Loligo. J. Physiol. **116**, 473–496 (1952).

[Hoese et al. 2003] Hoese W. J., Podos J., Boetticher N. C. and Nowicki S. Vocal tract function in birdsong production: experimental manipulation of beak movements. J. Exp. Biol. **205**(11): 1833–1843 (2003).

[Hoppensteadt and Izhikevich 1997] Hoppensteadt F. and Izhikevich E. M. *Weakly Connected Neural Networks*, Applied Mathematical Series, No. 126. Springer, New York (1997).

[Ishizaka et al. 1972] Ishizaka K. and Flanagan J. L. (1972). Synthesis of voiced sounds from a two-mass model of the vocal chords. Bell. Syst. Tech. J. **51**, 1233–1268 (1972).

[Izhikevich 2005] Izhikevich E. M. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*, in preparation (2005).

[Keener and Sneyd 1998] Keener J. and Sneyd J. *Mathematical Physiology*. Springer, New York (1998).

[Kimpo et al. 2003] Kimpo R. R., Theunissen F. E. and Doupe A. Propagation of correlated activity through multiple stages of a neural circuit. J. Neurosci. **23**, 5750–5761 (2003).

[Kinsler et al. 1982] Kinsler L., Frey A. R., Coppens A. B. and Sanders J. V. *Fundamentals of Acoustics*. Wiley, New York (1982).

[Kittel et al. 1965] Kittel C., Knight W. D. and Ruderman M. A. *Mechanics, The Berkeley Physics Course I*. McGraw-Hill, New York (1965).

[Koch 1999] Koch C. *Biophysics of Computation: Information Processing in Single Neurons*. Oxford University Press, New York (1999).

[Konishi 1965] Konishi M. The role of auditory feedback in the control of vocalization in the white crowned sparrow. Z. Tierpsychol. **22**, 770–783 (1965).

[Konishi 1994] Konishi M. An outline of recent advances in birdsong neurobiology. Brain Behav. Evolut. **44**(4–5), 279–285 (1994).

[Kroodsma and Konishi 1991] Kroodsma D. E. and Konishi M. A suboscine bird (eastern phoebe, *Sayonrnis phoebe*) develops normal song without auditory feedback. Anim. Behav. **42**, 477–487 (1991).

[Laje et al. 2002] Laje R., Gardner T. J. and Mindlin G. B. Neuromuscular control of vocalizations in birdsong: a model. Phys. Rev. E **65**, 051921 (2002).

[Laje and Mindlin 2002] Laje R. and Mindlin G. B. Diversity within birdsong. Phys. Rev. Lett. **89**, 288102 (2002).

[Laje and Mindlin 2003] Laje R. and Mindlin G. B. Highly structured duets in the song of the Southamerican hornero. Phys. Rev. Lett. **91**, 258104 (2003).

[Laje et al. 2001] Laje R., Gardner T. J. and Mindlin G. B. Continuous model for vocal fold oscillations to study the effect of feedback. Phys. Rev. E **64**, 056201 (2001).

[Landau and Lifshitz 1991] Landau L. D. and Lifshitz E. M. (1991). *Mecanica de fluidos*. Editorial Reverte, Barcelona (1991).

[Larsen and Goller 2002] Larsen O. N. and Goller F. Direct observation of syringeal muscle function in songbirds and a parrot. J. Exp. Biol. **205**, 25–35 (2002).

[Marler 1970] Marler P. A comparative approach to vocal learning: song development in white-crowned sparrows. J. Comp. Physiol. Psychol. **71**, 1–25 (1970).

[McCasland 1987] McCasland J. S. Neuronal control of birdsong production. J. Neurosci. **7**, 23–39 (1987).

[Mende et al. 1990] Mende W., Herzel H. and Wermke K. Bifurcations and chaos in newborn cries. Phys. Lett. A **145**, 418–424 (1990).

[Mindlin et al. 2003] Mindlin G. B., Gardner T. J., Goller F. and Suthers R. Experimental validation of a physical model for birdsong. Phys. Rev. E, 68, art 041908 (2003).

[Nordeen and Nordeen 1988] Nordeen K. W. and Nordeen E. J. Projection neurons within a vocal motor pathway are born during song learning in zebra finches. Nature **334**, 149–151 (1988).

[Nordeen and Nordeen 1997] Nordeen K. W. and Nordeen E. J. (1997) Anatomical and synaptic substrates for avian song learning. J. Neurobiol. **33**, 532–548 (1997).

[Nottebohm 1970] Nottebohm F. Ontogeny of birdsong. Science **167**, 950–956 (1970).

[Nottebohm 1976] Nottebohm F. Phonation in the orange-winged Amazon parrot, Amazona amazonica. J. Comp. Physiol. **108**, 157–170 (1976).

[Nottebohm 2002a] Nottebohm F. Why are some neurons replaced in adult brain? J. Neurosci. **22**, 624–628 (2002).

[Nottebohm 2002b] Nottebohm F. Birdsong's clockwork. Nature Neurosci. **5**, 925–926 (2002).

[Nottebohm et al. 1976] Nottebohm F., Stokes T. M. and Leonard C. M. Central control of song in canary *Serinus canarius*. J. Comp. Neurol. **165**, 457–486 (1976).

[Nottebohm et al. 1986] Nottebohm F., Nottebohm M. E. and Crane, L. Developmental and seasonal changes in canary song and their relation to changes in the anatomy of song control nuclei. Behav. Neural Biol. **46**, 445–471 (1986).

[Nowicki 1987] Nowicki S. Vocal tract resonances in oscine songbird production: evidence from birdsongs in a helium atmosphere. Nature **325**, 53–55 (1987).

[Nowicki 1997] Nowicki S. Bird acoustics. In *Encyclopedia of Acoustics*, ed. by Crocker M. J. Wiley, (1997), pp. 1813–1817.

[Nowicki and Capranica 1986] Nowicki S. and Capranica R. R. Bilateral syringeal interaction in vocal production of an oscine bird sound. Science **231**, 1297–1299 (1986).

[Podos 1996] Podos J. Motor constraints on vocal development in a song bird. Anim. Behav. **51**, 1061–1070 (1996)

[Press et al. 1992] Press W. H., Teukolsky S. A., Vetterling W. T. and Flannery B. P. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, New York, NY 2nd edition (1992).

[Reuter et al. 1999] Reuter R., Orglmeister R. and Herzel H. Simulations of vocal fold vibrations with an analog circuit. International Journal of Bifurcations and Chaos **9**, 1075–1088 (1999).

[Robb 1988] Robb J. B. and Saxman J. Acoustic observations in young children's vocalizations. J. Acoust. Soc. Am. **83**, 1876–1882 (1988).

[Scharff and Nottebohm 1991] Scharff C. and Nottebohm F. A comparative study of the behavioural deficits following lesions in various part of the zebra finch song system: implications for vocal learning. J. Neurosci. **11**, 2896–2913 (1991).

[Schuster et al. 1990] Schuster H. G. and Wagner P. A model for neural oscillators in the visual cortex: 1. Mean-field theory and derivation of phase equations. Biol. Cybern. **64**, 77–82 (1990).

[Sigman and Mindlin 2000] Sigman M. and Mindlin B. G. Dynamics of three coupled excitable cells with D_3 symmetry. Int. J. Bif. Chaos **10**, 1709–1728 (2000).

[Solari et al. 1996] Solari H. G., Natiello M. and Mindlin G. B. *Nonlinear Dynamics: a Two Way Trip from Physics to Math.* IOP, London (1996).

[Spiro et al. 1999] Spiro J. E., Dalva M. B. and Mooney R. (1999) Long-range inhibition within the zebra finch song nucleus RA can coordinate the firing of multiple projection neurons. J. Neurophysiol. **81**, 3007–3020 (1999).

[Stark and Perkel 1999] Stark L. L. and Perkel D. J. Two stage, input specific synaptic maturation in a nucleus essential for vocal production in the zebra finch. J. Neurosci. **19**, 9107–9116 (1999).

[Steinecke and Herzel 1995] Steinecke I. and Herzel H. Bifurcations in an asymmetric vocal fold-model. J. Acoust. Soc. Am. **97** 1874–1884 (1995).

[Straneck 1990a] Straneck R. *Canto de las Aves Pampeanas I*, LOLA, Buenos Aires (1990).

[Straneck 1990b] Straneck R. *Canto de las Aves Misioneras I*, LOLA, Buenos Aires (1990).

[Straneck 1990c] Straneck R., *Canto de las Aves Misioneras II*, LOLA, Buenos Aires (1990).

[Sturdy et al. 2003] Sturdy C. B., Wild J. M. and Mooney R. Respiratory and telencephalic modulation of vocal motor neurons in the zebra finch. J. Neurosci. **23**, 1072–1086 (2003).

[Suthers 1990] Suthers R. Contributions to birdsong from the left and right sides of the intact syrinx. Nature **347**, 473–477 (1990).

[Suthers 1994] Suthers R. A. Variable asymmetry and resonance in the avian vocal tract: a structural basis for individually distinct vocalizations. J. Comp. Physiol. A **175**, 457–466 (1994).

[Suthers 2001] Suthers R. A. Peripheral vocal mechanisms in birds: are songbirds special? Netherlands J. Zool. **51**, 217–242 (2001).

[Suthers 2004] Suthers R. A. The vocal apparatus. Ann. N.Y. Acad. Sci. **1016** 109–129 (2004).

[Suthers and Margoliash 2002] Suthers R. A. and Margoliash D. Curr. Opin. Neurobiol. **12**, 684–690 (2002).

[Suthers and Zuo 1991] Suthers R. A. and Zuo M. X. A test of the aerodynamics whistle hypothesis. Soc. Neurosci. Abstr. **17**, 1050 (1991).

[Suthers et al. 1999] Suthers R. A., Goller F. and Pytte C. The neuromuscular control of birdsong. Phil. Trans. R. Soc. London B **354**, 927–939 (1999).

[Suthers et al. 2002] Suthers R. A., Goller F. and Wild M. J. Somatosensory feedback modulates respiratory motor program of crystallized birdsong. Proc. Natl. Acad. Sci USA 99: 5680–5685 (2002).

[Szucs et al. 2000] Szucs A., Verona P., Volkovskii A. R., Abarbanel H. D. I. A., Rabinovich M. and Selverston A. I. Interacting biological and electronic neurons generate realistic oscillatory rhythms. Comp. Neurosci. Neuro. Rep. **II**, art. number 9278 (2000).

[Tchernikovski et al. 2001] Tchernikovski O., Mitra P. P., Lints T. and Nottebohm F. Dynamics of the vocal imitation process: how a zebra finch learns its song. Science **291**, 2564–2569 (2001).

[Thorpe 1961] Thorpe W. H. *Bird-Song*. Cambridge University Press, Cambridge (1961).

[Titze 1988] Titze I. R. The physics of small-amplitude oscillation of the vocal folds. J. Acoust. Soc. Am. **83**, 1536–1550 (1988).

[Titze 1994] Titze I. R. *Principles of Voice Production*. Prentice Hall, Englewood Cliffs, NJ (1994).

[Trevisan et al. 2005] Trevisan M., Mindlin G. B. and Goller F. A nonlinear model for diverse respiratory patterns of birdsong. Preprint (2005).

[Troyer and Doupe 2000a] Troyer T. W. and Doupe A. J. An associational model of birdsong sensory motor learning I. Efference copy and the learning of song syllables. J. Neurophysiol. **84**, 1204–1223 (2000).

[Troyer and Doupe 2000b] Troyer T. and Doupe A. J. An associational model of birdsong sensorymotor learning II. Temporal hierarchies and the learning of song sequence. J. Neurophysiol. **84**, 1224–1239 (2000).

[Vicario 1991] Vicario D. Contributions of syringeal muscles to respiration and vocalization in the zebra finch. J. Neurobiol. **22**, 63–73 (1991).

[Wild 1993] Wild J. M. Descending projections of the songbird nucleus robustus archistriatalis. J. Comp. Neurol. **338**, 225–241 (1993).

[Wild 1997] Wild J. M. Neural pathways for the control of birdsong production. J. Neurobiol. **33**, 653–670 (1997).

[Wild 2004] Wild J. M. Functional neuroanatomy of the sensorimotor control of singing. Ann. N.Y. Acad. Sci. **1016**, 1–25 (2004).

[Wild et al. 1998] Wild J. M., Goller F. Suthers R. A. Inspiratory muscle activity during birdsong. J. Neurobiol. **36**, 441–453 (1998).

[Wilden et al. 1998] Wilden I., Herzel H., Peters G. and Tembock G. Subharmonics, biphonation and deterministic chaos in mammal vocalization. Bioacoustics **8**, 1–30 (1998).

[Yu and Margoliash 1996] Yu A. C. and Margoliash D. Temporal hierarchical control of singing in birds. Science **273**, 1871–1875 (1996).

[Zollinger and Suthers 2004] Zollinger S. A. and Suthers R. A. Motor mechanisms of a vocal mimic: implications for birdsong production. Proc. R. Soc. London B **271**, 483–491 (2004).