CHAPTER
1

Nonlinearity, Complexity, and Control in Vocal Systems

Neville H. Fletcher

Our emphasis in this book is on the complexity of the behavior of the human vocal system. It is trite but true to say that all biological systems are complex and nonlinear in their response—greater stimulus does not simply increase the magnitude of the response but may bring about a qualitative change in its nature. Here we focus on just a small part of that nonlinearity by considering the physical behavior of the vocal folds and the mechanisms and strategies by which their complex behavior can be controlled.

In the case of the human vocal system, we encounter nonlinearity in the vocal fold vibrations themselves, in the aerodynamics of the flow through the vocal fold opening, and in the interaction between these two quantities. This highly nonlinear active part of the vocal system is coupled to passive and very nearly linear multimode vocal tract resonators both downstream and upstream from the glottis, and these interact with the motion and flow mechanics of the vocal folds. While it would seem that the nonlinearity and physical complexity of such a system might make its control a matter of great difficulty, it turns out, paradoxically, that the nonlinearity itself tends to lock the system into stable regimes of oscillation that can be brought under coordinated neural control.

Human speech or song is a combination of phase-locked, harmonic, voiced vowel sounds, and near-chaotic unvoiced stops and consonants, produced in rapid time-sequence. The same is true of the sounds produced by many other animals, although there are cases in which the vocal sound is both voiced and chaotic at the same time. The complexity of this behavior presents a rich field for study.

VOCAL FOLD OSCILLATIONS

The human glottis contains two approximately symmetrical vocal folds which can move to control the air flow through the larynx. Together they form a pressure-controlled valve that can be set into autonomous oscillation through interaction with the acoustic pressures in the respiratory tract above and below the larynx. Simple single-mass models [1,2] exhibit most of the characteristics of such systems for various possible directions of motion of the valve flaps, but realistic models for the human vocal folds must contain many more configurational parameters because of the flexible nature of biological tissue. The two-mass model of Ishizaki and Flanagan [3] is a well-known example of an extension in this direction, though much more complex models have also been developed. Because of the nonlinearities we discuss in the following, it is always necessary to resort to numerical methods to calculate the acoustic behavior of these system models [2,3].

The exact geometry of flow separation at the vocal folds is important to details of the valve operation and has usually been treated simply by assumption. The acoustic impedance of the air passages below the larynx, terminating in the lungs, has surprisingly also been neglected in many models, but it can have a quite significant influence on behavior. Indeed the conditions for autonomous oscillation of the vocal valve involve not only the properties of flow through the valve itself, but also the acoustic impedances of the ducts leading to and from it [1]—the bronchi and lungs upstream and the trachea and mouth downstream of the glottis. Valve oscillation is maintained by a combination of the influences of fluctuating pressure in the regions above and, more importantly, below the glottis, together with pressure differentials caused by Bernoulli effects in the flow through the valve aperture. Which driving mechanism is more important depends upon details of the valve geometry and of the separation of flow though it, and indeed we might expect these details to change progressively throughout the vocal range in the case of singers.

Note that, although there is some similarity between the operation of the vocal folds and that of the lips when playing a brass instrument such as the trumpet, there is also a very significant difference. In the vocal fold case, the frequency of vibration, which is controlled by the mass and muscular tension of the vocal cords, is much less than that of the first resonance of the vocal tract, except for high notes of a soprano voice. When playing a brass instrument, on the other hand, the natural frequency of the lip vibration is adjusted so as to nearly coincide with the frequency of one of the resonances of the instrument horn. This difference has important consequences for the exact physics of the oscillation.

One important point, usually ignored, is that the two vocal folds will generally differ in mass and tension and thus in natural resonance frequency. The motion of the folds can, however, be separated into two normal modes with slightly different frequencies: one in which the two folds move cophase relative to the symmetry plane and therefore close at one extremity of their motion, and one in which they are antiphase, as shown in Figure 1–1. The antiphase motion does not change the flow

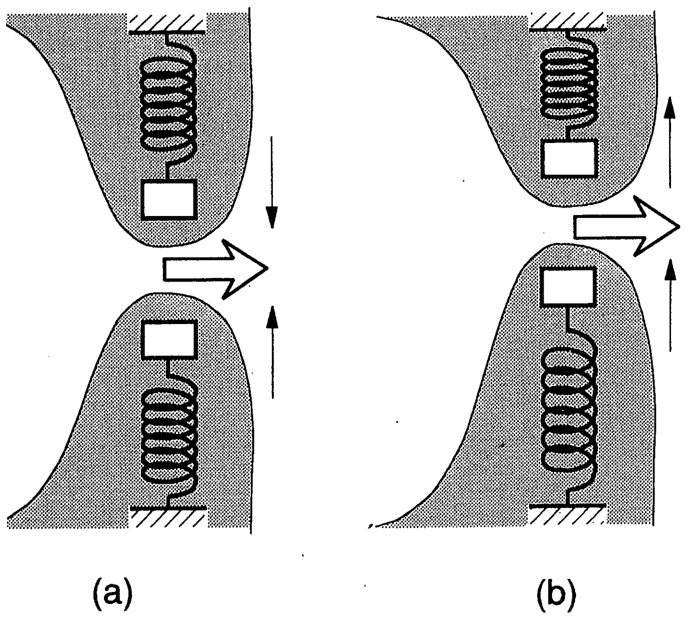

Figure 1–1. **(a)** Symmetric or cophase oscillation; and **(b)** antisymmetric or antiphase oscillation of vocal folds. The underlying model indicated is a "single-mass" model.

through the glottis, so that it cannot be maintained by feedback effects. The cophase motion, however, couples directly to the flow and can be maintained by appropriate feedback from flow-generated pressures. In a real vocal system with appreciable damping, we can therefore usually ignore the antiphase motion and treat the symmetrical cophase motion as the only mode of significance. This is a completely linear effect.

In a more realistic model, and particularly as we approach the reality of a model with a continuous distribution of mass and elastic properties, there are of course many other vibrational modes that should be considered. In the simple two-mass model [3], the larynx is considered to be two-dimensional, with each vocal fold consisting of a pair of masses, one on the leading and one on the trailing edge of the fold, appropriately coupled together by springs as shown in Figure 1–2(a). As is well known, numerical calculations with this model show that the mode of vocal interest is one in which there is a phase difference between the leading and trailing masses or, expressed in another way, a standing

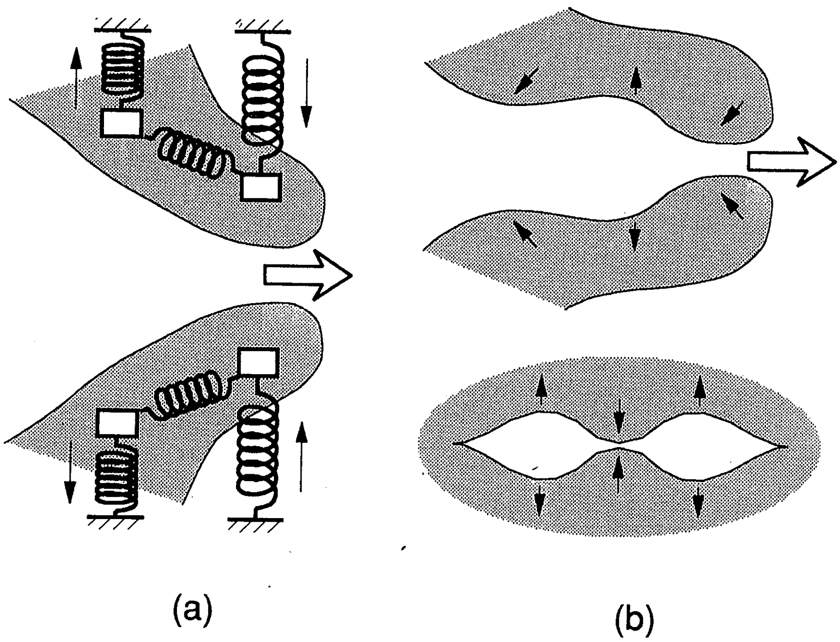

Figure 1–2. **(a)** Phase differences in a simple two-mass model; **(b)** some possible higher vibration modes in a more realistic continuum model. In (b) the upper drawing is a cross section and the lower drawing a face-on view of the glottis.

wave along each vocal fold, in the air flow direction, synchronized with the basic opening and closing mode for the fold pair as shown in Figure 1–2(a). The nonlinearity of the system, which we discuss in a moment, is important in synchronizing these two vibration modes to give a well organized and apparently simple vocal fold oscillation.

In a fully three-dimensional model, which comes closer to physical reality, we must allow for the possibility of phase differences in the motion of the folds, not only parallel to the air flow but also in directions normal to it, as shown in Figure 1–2(b). Each fold, indeed, is able to behave to some extent like a thick elastic plate, though this analogy is of little assistance because of the complex shape of the folds. The various possible vibrational modes of the folds have frequencies that are completely unrelated. Some of these vibrations can be ruled out of consideration in a perfect larynx because their symmetry is such that they do not affect the glottal flow, although the asymmetries of a deformed or damaged vocal fold may reintroduce this coupling.

The principal sources of nonlinearity in the vocal fold oscillation are illustrated in Figure 1–3. In a linear system, the effective vibrating mass, the spring constant, and the damping coefficient are all constant over the oscillation cycle, and the resulting differential equation describing the

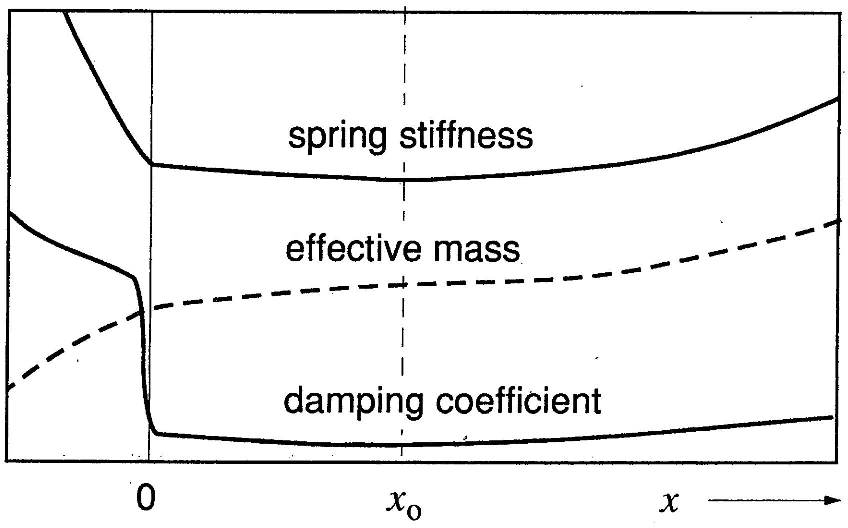

Figure 1–3. Qualitative behavior of the spring constant and the damping coefficient for vocal fold oscillation. The coordinate x measures the opening between the folds, with x_0 being the equilibrium opening distance.

motion is linear, in a mathematical sense. For vocal fold oscillations this is true for very small motions about the equilibrium position x_0 but the approximation breaks down for larger oscillations. As the vocal folds open more widely, the spring stiffness of the biological tissues increases and so does the effective mass recruited into the vibration. As the vocal folds close together, however, the departure of these parameters from constancy becomes extreme—the spring stiffness increases greatly and so does the damping coefficient, because the tissues of the two folds are being distorted by their contact. These nonlinearities are enhanced if the equilibrium opening distance x_0 is small and if the amplitude of the vocal fold oscillations is large.

The effects of nonlinearity upon oscillatory behavior are many and varied. If we consider just a single mode of vibration, for example in a single-mass model, then the effect is simply to distort the normal sinusoidal time variation of the opening dimension x so that the waveform is clipped softly at extreme openings and sharply when the folds close together. This introduces phase-locked harmonics of the fundamental in the motion and hence in the air flow through the glottis. The strength of these harmonic partials increases with the amplitude of the vocal fold vibration; initially the amplitude of the nth harmonic increases as the nth power of the amplitude of the fundamental. This is reflected in the air flow through the glottis and in the sound quality of the voice, which progresses from dull to strident as the subglottal air pressure, which controls vocal fold vibration amplitude, is increased.

In the case of a multimass model of the vocal folds, and in particular for the realistic case in which the mass distribution is continuous, there are many possible vibration modes with frequencies that have no simple numerical relation between them. In is a characteristic of such multimode systems that a sufficient degree of nonlinearity can lock all these vibrations into exact harmonic frequency relationship, so that the resulting motion is simple and periodic [4]. This almost always happens in vocal fold oscillations, because the closing nonlinearity is extreme, and to it we owe the fact that human vowel sounds, and indeed most voiced animal sounds, have strictly harmonic upper partials, as can be seen from the fact that the waveform of the steady sound repeats exactly in each cycle.

As well as the mode-locking behavior that can be induced by strong nonlinearity, nonlinear systems are also capable of a variety of complex behavior, depending upon details of their excitation. If the system is sufficiently simple, such as a single-mass model, then we can find the normal progression from a period-doubling bifurcation, giving a subharmonic an octave below the normal frequency, through further successive bifurcations to chaotic behavior [5]. This chaotic behavior is called deterministic chaos, because it is governed by simple underlying laws, and the tools that have been developed for its study are many. In particular, we can derive much of the detail from a consideration of the time series representing the oscillation [6] and from that a strange attractor with fractal geometry. Needless to say, if the vocal fold oscillation is bifurcated or chaotic, so too will be the flow through the glottis and hence the vocal sound. We return to this later.

Any realistic model of the human vocal folds necessarily involves a description of its configuration involving more than a single displacement parameter—the dimensionality of the system is greater than unity. Such complex systems can again show two sorts of behavior: Either they can become locked into a stable oscillation regime or they can exhibit bifurcations and chaos. The stable regime is simple to treat, since the variations of all coordinates are harmonically related, but the chaotic regime can be very complicated. It is not possible, in general, to derive simple attractor mappings from the time series representing the oscillation, because its description may involve a phase space of many dimensions. Any projection onto a space of lower dimensionality then results in a featureless cloud from which nothing can be deduced.

While examples of chaotic vocal fold oscillation are happily rare in the case of humans, the same is not necessarily true of other animals. An excellent example is provided by the Australian sulfur-crested cockatoo (Cacatua galerita). The syrinx of a bird is in many ways analogous to the larynx of a human, but there are several distinct differences. In the first place, the valve in the syrinx is not a single structure lying in the trachea above the junction of the two bronchi but rather a pair of independent valves, one in each bronchus, lying just below the junction with the trachea. In the second place, each valve consists of a membrane forced into the airway by pressure in a surrounding air sac, rather than a flap of tissue tensed by cartilage. Nevertheless, the operation of the syrinx can be described by a simple nonlinear model that is quite similar to a single-mass model for the human vocal folds [7]. While mostly exhibiting a simple phase-locked oscillation regime similar to that of the human vocal folds, this model can be induced to display complex and chaotic behavior. It is to be expected that a more realistic model, with the syrinx membranes described by several geometric parameters, would behave similarly. The cockatoo Cacatua galerita is one of those birds that can be trained to imitate human speech, but its natural call is an extremely loud and raucous screech with a wide-band spectrum peaking around 7 kHz. Analysis of the resulting time series suggests that the dimensionality of the oscillation is high, reflecting a multiparameter oscillating system, so that it is difficult to treat. We should perhaps be grateful that the normal speech of humans is more disciplined!

FLOW NONLINEARITY

Oscillation of the vocal folds, as we have seen, is nonlinear, with the folds spending an appreciable part of each cycle in the closed state and with the extent of peak opening limited by elastic nonlinearities. The extent of this oscillation nonlinearity will increase as the energy of the oscillation increases, an effect that can be achieved by increasing the pressure below the glottis. Muscular control of the static separation of the vocal folds can also influence the fraction of each period spent in the closed state, although under nearly all assumptions the vocal folds do close once in each cycle.

The source spectrum for the voice is determined by the flow of air through the glottis, and therefore by the product of the area of the glottal opening and the velocity of air flow. The open cross section of the glottis does not behave in a simple sinusoidal fashion, as we have seen, because of elastic nonlinearities and collision of the vocal folds upon closure. The extent of this departure from sinusoidal behavior is at least partially under muscular control. Flow through the vocal fold aperture is itself nonlinear and governed to a good approximation by the Bernoulli equation, which shows that flow velocity is proportional to the square root of the pressure drop across the aperture. This Bernoulli flow nonlinearity is convolved with the nonsinusoidal variation of the aperture area and with the variation of pressure below and above the vocal folds and further distorts the flow waveform. Both the acoustic power in the glottal source and the fraction of energy in high harmonics are increased if the subglottal pressure is high, though the conversion efficiency from pneumatic energy in flow from the lungs to acoustic energy in the vocal tract is typically less than 10%.

As is well known, the flow spectrum is then multiplied by the resonant response of the vocal tract to impress upon it the formants that are a vital part of speech. In principle, there are also formants that include the part of the vocal tract below the glottis during the part of each cycle in which the vocal folds are open. The prominence of these formants depends upon the damping of the complete vocal tract, including its lower termination in the lungs. It seems that this termination is sufficiently resistive in humans that these full-length formants are not observed, though the opposite appears to be true of at least some birds [7]. We should also, in principle, take into account the interaction of the high-frequency pressure fluctuations in the trachea upon the motion of the vocal folds. In practice, however, since these formant oscillations are at frequencies that are a high multiple of the vocal fold vibration frequency, they have little effect. They are important, however, in influencing the instantaneous flow through the open glottis.

Because of flow losses in the glottis and viscous and thermal losses in the vocal tract, the total conversion efficiency from pneumatic energy supplied by the lungs to acoustic energy radiated from the mouth rarely exceeds 1%, and is often much lower [7]. Similar figures apply to sound-producing systems of other types, including musical instruments and electromagnetic loudspeakers.

COMPLEX FLOW BEHAVIOR

Complexity of flow behavior, apart from that introduced by complex behavior of the vocal folds, can arise in two ways. The first, which can occur even in laminar flow, is a variation of the point of detachment of the flow from the trailing edge of the valve aperture, as shown in Figure 1–4. Such a variation can clearly have a large effect on the Bernoulli pres-

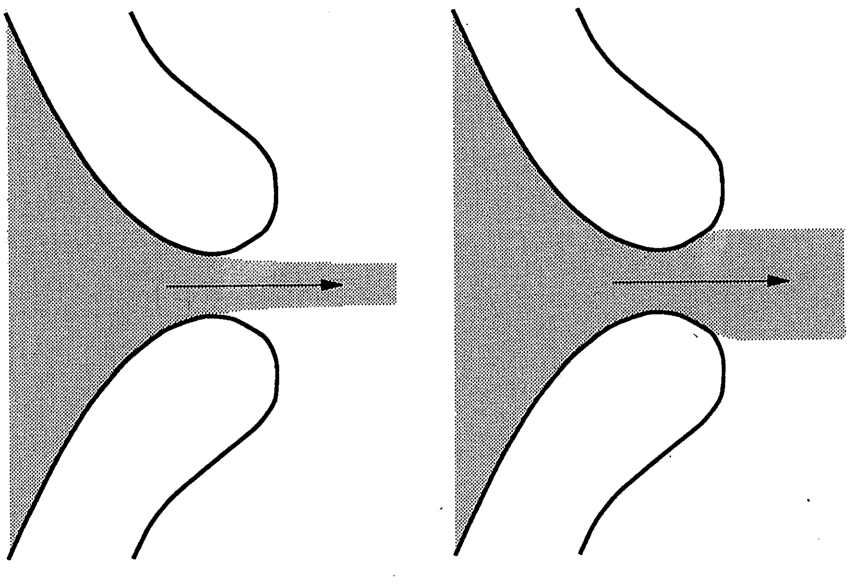

Figure 1–4. Variation in flow detachment points. **(a)** Early flow detachment with a vena contracta effect that further reduces the jet cross section; **(b)** late detachment with partial expansion of the jet after the constriction.

sures acting on various parts of the vocal folds and thus upon their motion. If this variation is periodic then it simply adds to the harmonic development, but more complex behavior is possible. Because the process is nonlinear it can exhibit period bifurcation and thus the generation of subharmonics. In the limit this can lead to chaotic behavior and turbulence. Since flow behavior is responsible for acoustic output and also reacts back upon the motion of the vocal folds, these complexities of flow are important for understanding voice quality.

The second effect, which is largely responsible for the wide-band noise associated with fricatives and sibilants, derives from turbulent flow either at the larynx or in the mouth. This turbulence generally has so many possible degrees of freedom that it is truly chaotic rather than deterministically chaotic, as we find with systems of lower dimensionality. There may, however, be deterministic elements embedded in the flow and associated with quasiregular vortex shedding from parts of the bounding structure. It is difficult to do much with turbulence except to describe its fluctuation spectrum and associated radiated acoustic power.

CONTROL

When controlling a linear system, one has simply to set the values of controlling parameters, such as muscular tensions and blowing pressures, and then allow sufficient time for the system to settle into equilibrium. This time is controlled by the system Q value and is typically a few tens of cycles of the fundamental oscillation. In the control of a nonlinear system, on the other hand, we must pay attention not just to the final state but also to the trajectory in parameter space by which that state is approached, since the final oscillation regime can be a function of system history.

This complexity might at first sight appear quite daunting, but in fact it can have some advantages from the point of view of control, particularly when we wish to produce several different kinds of output. A deterministically chaotic system has two important properties. The first is that its trajectory in phase space goes close to every possible point on the attractor and thus to every possible regime of oscillation, usually within rather few cycles of the basic oscillation. The second is that the future behavior of the trajectory, and thus the form of the future oscillation, can be influenced to an extremely large extent by a very small change in the parameters of the system. Together these two facts mean that any desired mode can be stabilized, at least in principle, by repeated application of very small corrections to the system parameters. They also mean that, if we wish to change the regime of oscillation, then we just need to let the system run autonomously for a short time until its trajectory brings it close to a point on the orbit we wish to choose, when a small correction can stabilize this new regime.

When such a control operation is considered ab initio, its possible complexity is daunting. It is quite another thing, however, when the control is exercised by a human neural system. Such a neural system behaves as an elaboration of the simple neural networks that are now exciting increasing attention as intelligent decision and control systems. A simplified example is shown in Figure 1–5. Each element of the system mimics a neuron, in that it has several input connections and a single output, and

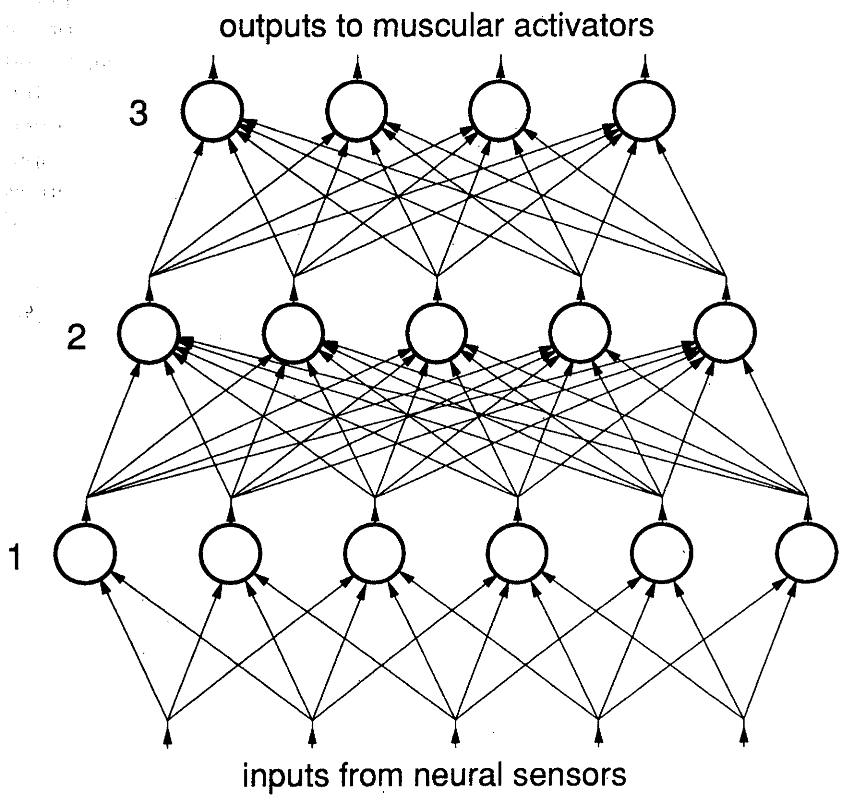

Figure 1–5. A simple neural network. Layer 1 of the network accepts stimuli from neural sensors in various parts of the body, while layer 3 produces control signals for the vocal muscles. Layer 2 is a hidden layer with no external connections. The arrows indicate signal-path connections between the three layers, each connection having an assigned weight that can be changed during the training process.

has only two output states, "on" and "off." Whether a neuron is in the on or the off state is determined by a weighted average of the on or off signals applied to its inputs, the weighting factors for the input links being important parameters of the network. The elements are arranged in at least three layers and there is a great deal of interconnectivity between them. The neurons of the lowest layer sense the physical state of the system at some instant and each turns on or off depending upon the stimulus it receives. The outputs of this layer feed to the second "hidden" layer and turn its elements on or off depending upon the weightings assigned to the network links. The outputs of the elements of the second layer now pass on to the third layer and similarly determine its pattern of on and off states. There is thus a wave of "neural" activity that passes through the system from its first to its third layer, and we can suppose the process to be repeated many times each second as the state of the system changes. The third layer is the output layer of the network, and its pattern of on/off states directly controls muscular actuators that influence the state of the system. Note that the elements of the neural network are highly nonlinear in their response—a small change in the input states can flip a neuron from its off state to its on state. The network as a whole is nonlinear in a very much more subtle and complex way, and its outputs are not simple functions of its inputs.

We can represent this more formally by supposing that O_i is the output signal, either 0 for "off" or +1 for "on," from the ith neuron. The input signal I_j to the jth neuron is then

$$I_j = \sum_i w_{ij} O_j$$

where the w_{ij} are the synaptic weighting factors, either positive or negative, for the links between neurons i and j. Neuron j will now fire and produce an output signal of +1 if $I_j > T$ where T is a threshold value built into the neuron. Each neuron of the network has a similar behavior, but the weighting factors w_{ij} are different for each link of the network.

Neural networks "learn" by modifying the weights assigned to their internal synaptic connections so as to reinforce desired outputs and inhibit those that are undesired. In this way they can maintain a system in a desired state, which may be either constant or else varying in time in a regular manner. The same approach can be applied to system trajectories. The time-varying weighting parameters corresponding to the desired trajectory are stored somewhere outside the network and then passed to it in sequence. This is quite clearly closely related to the way in which humans learn complex tasks—first we learn to maintain steady states such as vowel sounds and then to combine them to produce desired outputs such as phonemes and then words. One might speculate that the problems associated with a "breaking" voice in adolescent male humans are not so much due to physical changes in the larynx, although these are certainly present, but rather to a time lag in adjusting the synaptic weights in the associated neural control system.

With simple neural networks, and even more strikingly with real neural control systems, it is difficult or even impossible to analyze the behavior at a component level—the system is not a conventional von Neumann machine, like an ordinary computer, but something altogether more complex.

CONCLUSIONS

While the physiology and physics of voice production are both interesting subjects in their own right, the importance of studying these matters comes from the fact that vocal communication is a vital part of our everyday life and cultural heritage. The more we can understand about the complexities of the vocal system and its control mechanisms, the better we will be able to repair its defects and to learn or teach about its exploitation.

This book focuses on nonlinearity and complexity in both the vocal system and its neural control. As we have explained above, the nonlinearity of the system is responsible for the richness of vocal sounds, and it is also responsible for the fact that these sounds generally have regularly repeating waveforms and thus harmonic spectra. The nonlinearity locks the behavior of the system into one of a number of regimes of autonomous oscillation, which can then be controlled as stable entities. At the same time, nonlinearity allows the possibility of more complex and even chaotic behavior, and it is generally the objective of the neural control system to avoid these.

A recognition of the nature of the neural control system, and its simplified modeling as a neural network, shows us the role of both imitative practice and conscious thought in achieving both routine vocal utterances and special vocal effects. Certainly this approach has been exploited by evolutionary processes in natural childhood development, and recognized by teachers for millenia. A more thorough understanding of the science upon which these methods are based should make them even more effective.

REFERENCES

1. N.H. Fletcher, "Autonomous vibration of simple pressure-controlled valves in gas flows," *J. Acoust. Soc. Am.* **93**, 2172–2180 (1993).
2. S. Adachi and M. Sato, "Time-domain simulation of sound production in the brass instrument," *J. Acoust. Soc. Am.* **97**, 3850–3861 (1995).
3. K. Ishizaka and J.L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell Syst. Tech. J.* **51**, 1233–1268 (1972).
4. N.H. Fletcher, "Mode locking in nonlinearly excited inharmonic musical oscillators," *J. Acoust. Soc. Am.* **64**, 1566–1569 (1978).
5. P. Cvitanovic (Ed.), *Universality in Chaos*, (Adam Hilger, Bristol 1984).
6. N.H. Packard, J.P Crutchfield, J.P Farmer and R.S. Shaw, "Geometry from a time series," *Phys. Rev. Lett.* **45**, 712–716 (1980).
7. N.H. Fletcher, *Acoustic Systems in Biology*, (Oxford University Press, New York 1992) Ch. 14.