Categorization of Musical Patterns by Self-Organizing Neuronlike Networks

ROBERT O. GJERDINGEN
SUNY at Stony Brook

Simulations of self-organizing neuronlike networks are used to demonstrate how untrained listeners might be able to sort their perceptions of dozens of diverse musical features into stable, meaningful schemata. A presentation is first made of the salient characteristics of such networks, especially the adaptive-resonance-theory (ART) networks proposed by Stephen Grossberg. Then a discussion follows of how a computer simulation of a four-level ART network—a simulation dubbed L’ART pour l’art—indepedently categorized musical events in Mozart’s six earliest compositions. The ability of the network to abstract significant voice-leading combinations from these pieces (and in fact to detect a possible error in the New Mozart Edition) suggests that this approach holds promise for the study of how ordinary listeners process music’s multidimensional complexity. In addition, the categorizations produced by the network are suggestive of alternative conceptualizations of music’s hierarchical structure.

In a prim household, tidiness is a paramount virtue. Dishes line cupboards like soldiers on parade. Books march across shelves in close-order drill. There is a place for everything, and everything is in its place. A visitor senses order, control, and the proud demonstration of rational forethought. Someone visiting the edifice of music theory would note that, here too, good housekeeping is prized. We value systematic order and structure to such a degree that our students must at times feel like employees in a professional cleaning service, one that operates in a particularly slovenly neighborhood. Again and again they are sent out to put stray pitches into chordal cupboards, to straighten out meandering melodies, to scrub away ambiguities, to do in fact whatever is necessary to leave the impression of thoroughgoing tidiness. But is music—and especially the perception of music—as tidy as our theories might wish it to be? Is it fair, for instance, to sweep non-harmonic tones under the rug or to paper over cracks and seams in the

Requests for reprints may be sent to Robert O. Gjerdingen, Department of Music, State University of New York at Stony Brook, Stony Brook, NY 11794-5475.

339
structural framework? Do we risk confusing the puritanical tidiness of logical thought with the sybaritic clutter of aesthetic experience?

In a recent issue of this journal, Bharucha (1987) presented a strong case for using computer simulations of neuronlike networks to model the perception of music. Neuronlike networks, termed “connectionist models” by Bharucha, Feldman (1982), and others, do have properties that make them well suited to emulate many of the complex, cluttered mental processes associated with listening to music. As great an authority as the human brain has found networks of neurons to be indispensable for the interpretation, storage, and retrieval of musical information. But as Carpenter (1989) reports in a review of neural-network models, there are major differences between the types of models currently proposed. Impressive results in simple, highly circumscribed domains may not be extensible to the perceptual complexities of real music.

A danger in modeling what Rumelhart and McClelland (1986) term “the microstructure of cognition” is that we may inadvertently impose our own predilections for neatness and theoretical parsimony on the incredibly complicated interactions that take place in a neuronlike network. For instance, when Bharucha chose to preassign his neuronlike processing units to specific pitches, chords, and keys, I believe he was forcing tonal phenomena into a hierarchy tidier than that afforded by musical experience. It is true, of course, that pitches, chords, and keys are empirically demonstrable facts of music perception. Moreover the nesting of pitches within chords within keys is a beautifully ordered structure, one of the crowning achievements of eighteenth-century music theory. Yet as a twentieth-century theorist, I do not feel at home in this ancien régime. For me, the structures of even so fastidious a composer as Mozart appear as dense tangles of contrapuntal lines, thorough-bass patterns, harmonic and melodic schemata, metrical frameworks, rhythmic gestures, and a host of more ineffable features.

A well-established but today perhaps less well publicized area of neural-network research focuses on so-called self-organizing networks—networks that determine their own categories and structures. The following discussion will explore how a four-level, self-organizing neuronlike network, one following the biologically oriented approach of Grossberg (1976, 1978, 1980, 1982) and his associates at Boston University’s Center for Adaptive Systems, can be used to simulate the formation of schemata for musical events and event sequences. What I will term Grossberg networks make important distinctions between the processes of neural excitation and inhibition, explicate the interplay between short-term and long-term memory, and restrict synaptic learning to real-time information available at the local synapse. These networks are capable of independently arriving at untidy but nevertheless quite interesting categorizations of musical events.
Level One: A Dynamic Short-Term Memory

Figure 1 shows an idealized musical short-term memory. The horizontal disk stands for a dense laminar field of interconnected neurons. The arrows rising from below this field indicate input sent from lower-level networks where acoustic information from the ear is processed into recognizable musical features. For the sake of illustration, let us imagine three abstract musical features, and let us depict those features by a triangle, a circle, and a square. The three peaks above the field stand for cells that have become especially active in response to the sounding of the three features. To understand the left-to-right order and relative intensities of these responses, we need first to elaborate on the notion of a musical short-term memory.

At the very least, a musical short-term memory must keep track of two different things: (1) the set of recent events and (2) their temporal order. The first requirement can be met by the selective excitation of individual cells, thus mapping a set of features onto a set of cells. The second can be met by enforcing a recency effect, whereby the strongest activation among a set of cells represents the trace of the most recently occurring feature. A direct procedure for achieving a recency effect uses a cell’s natural decay of activation in the absence of new excitation. If one knows the rate of that decay, then the residual level of activation provides a measure of how long it has been since a cell population last received excitation.

The use of decaying activations to record temporal order, while attractively simple, nonetheless has several drawbacks. A new but weak input signal might appear to be the decayed remnant of a strong signal, thus confus-

Fig. 1. A symbolic representation of three musical events (the abstract shapes). From the ear, signals are conveyed by the auditory system (the vertical arrows) to a laminar neural field (the horizontal disk), where these events stimulate responses in three neural populations.
ing past and present. Where cells specifically excite or inhibit each other, levels of activation will be altered in ways that can easily destroy any direct relationship between levels of activation and temporal order. Passive decay also allows the number of events stored in short-term memory to become unrealistically large, because a slow rate of decay, a large number of cells, and a rapid series of events could result in dozens if not hundreds of cells being simultaneously active at minutely different levels. Conversely, if cell populations decay passively without any reinforcement, an extended delay in the onset of a new event could allow short-term memory to go completely blank.

The problem hidden behind the use of passive decay for determining temporal order is that passive decay answers the clocklike question, “When did this event last occur?” when all that is needed is an answer to the queue-like question, “Before what event did this event occur?” A recency effect sufficient to encode temporal order can in fact be achieved without any reliance on decay. In place of the passive decay of individual cells, one substitutes the dynamic interaction of multiple groups of cells, each small group forming a feedback loop.

Three such interacting groups of idealized neurons are shown in Figure 2. An individual group consists of four cells (the white spheres), each exciting the next in counterclockwise rotation. A feedback loop is, of course, subject to a spiraling increase in excitation as each cell adds to the next’s level of activation. Without some type of gain control, each loop would soon saturate at its maximum level of activation. And with all three loops maximally active, no temporal ordering could be discerned. The network must therefore coordinate gain control across the entire field. One method, a pattern of so-called on-center, off-surround interactions, will be discussed later in the section “Level Two.” The method illustrated in Figure 2—one suggested by Carpenter and Grossberg (1987)—involves the use of inhibitory interneurons (the black spheres). An interneuron receives excitation from all the initial cells at one layer of the network \( (e_i, i = 1 \ldots n) \) and, if the magnitude of that summed excitation \( \langle M, \text{where } M = \langle \Sigma e_i \rangle \rangle \) exceeds a threshold \( T \), inhibits the next cells in the same layer by a factor of \( T/M \). A balance will be achieved between feedback excitation and interneural inhibition so that the maximum total activation passing between the lower and upper layers of the network stays very close to the value of \( T \), assuming as Grossberg does (1982, p. 222) that the two interneurons have similar thresholds. There is, in addition, a practical threshold to the minimum excitation that can circulate around each feedback loop. Carpenter (1989) notes that in this type of neural circuit, a nonlinear signal function (defining how a cell’s internal activation is transformed into an output signal) can have the effect of suppressing very weak signals, thus preventing the amplification of noise.
Fig. 2. A dynamic short-term memory. Nonspecific inhibitory interneurons (black spheres) control the gain of activation across the three feedback loops (white spheres). Arrows show excitatory connections, bars show inhibitory connections. Thick lines show the main paths of excitation. Thin lines show the ancillary paths used in controlling gain.

Because the total activation at each layer of the network is limited, a strong input to one feedback loop will decrease the activations in the other two loops. If the input is of any significant duration, the cells that it excites will quickly become the most highly active in the network. This creates the required recency effect. Referring back to Figure 1, the height of each peak above the field is drawn proportionally to each cell group’s level of activation. The peak for the third feature (the square) is highest because that feature was the one most recently heard. As the sequence would have been performed, each new feature in the present would strongly excite the appropriate groups of cells, which in turn would inhibit the cells representing features from the recent past. When the second feature (the circle) sounds, its cells inhibit the previously excited cells for the first feature (the triangle), and when the third feature (the square) sounds, its cells inhibit the cells for the second feature and further inhibit the cells for the first feature. The left-to-right order of cell groups thus has no special significance. Only the relative levels of activation encode the temporal order of events. Although music may move from left to right on the printed page, there is little likelihood that it does so in the brain.
An important property emerges from this type of neural architecture—that of a transient memory span (TMS). Because there is a limit to total activation, old activations must be reduced after the addition of each new activation. And because the nonlinear signal functions suppress weak signals, a past event can be inhibited by only so many new events before its trace disappears from the network. Simply put, new events crowd out the old. In the simulations described below, the TMS varies from four to six events, a range compatible with estimates of human short-term memory.

**Level Two: The Formation of Stable Categories**

For a network to retain a memory of a long series of events, a long melody for instance, it must in some way segment the series into parts short enough for each to fit within the limits of short-term memory. Grossberg networks accomplish this process of segmentation, often called chunking, by combining different neural fields in a multilevel hierarchy. Figure 3 shows two fields connected so that a single group of cells in the upper level is capable of categorizing an entire pattern of activation at the lower level. The lower field is the same as the one discussed earlier. Its pattern of activation is carried to upper-level groups of cells by axons represented symboli-
cally by the two heavy vertical lines linking the fields. The two fields are thoroughly interconnected, there being a path from every group of lower-level cells to every group of upper-level cells and vice versa. To avoid repeating such cumbersome locutions as “upper-level group of cells,” let us agree henceforth (following Grossberg) to abbreviate “lower-level field” as simply “F₁,” “upper-level field” as “F₂,” and “group of cells” as “population.”

Grossberg networks have no explicit inputs from an external, omniscient teacher. That is, unlike some neural-network models, a Grossberg network is not continually informed of the right answer to each problem of categorization. The network itself must decide which F₂ population is the right one to respond to each F₁ pattern of activation. And the network itself must decide whether an F₁ pattern of activation is a variant of an established category or a completely new pattern. For the purposes of modeling the cognition of musical phenomena, these are highly desirable traits. After all, the average listener receives no instruction in categorizing musical phenomena. And even trained musicians intuitively abstract and categorize most musical phenomena long before they begin the formal study of music theory. When it comes to the perception of music, we seem to lift ourselves by our own bootstraps.

Three mechanisms make possible the self-organizing formation of stable upper-level categories: (1) recurrent on-center, off-surround interactions, (2) modification of synaptic connections, and (3) stabilizing feedback in concert with an orienting subsystem. Let us examine each of these in turn.

**RECURRENT ON-CENTER, OFF-SURROUND INTERACTIONS**

Figure 4 shows a schematic close-up of the surface of the upper-level neural field (F₂). F₁ excitation reaching the spot marked with plus signs will initiate positive feedback within the local F₂ population. This is the “recurrent on-center.” The resulting activation in the on-center population has the effect of inhibiting the other F₂ populations, all marked with minus signs. This is the “off-surround.” Each F₂ population forms an on-center, with all other F₂ populations forming its off-surround. The entire field is thus a patchwork of on-centers with highly overlapping off-surrounds.

If we assume that the initial connections from F₁ to F₂ have random strengths, then any specific pattern of F₁ activation will most likely excite one F₂ population a little more than the others. All the populations may receive some excitation, but one will randomly receive the most. The F₂ population receiving the most excitation simultaneously strengthens its own activation (through internal feedback) and inhibits the activation of the other F₂ populations. Their activations will, of course, also inhibit the dominant population. But because on-center feedback and off-surround in-
hibitation increase quadratically with the amount of excitation received, the dominant population gains an ever-increasing advantage over the others. The end result will usually be one $F_2$ population with a high level of activation, possibly a few others with lower levels of activation, and the rest with no or even negative activation. In the simulations described below, the interactions were such that only a single $F_2$ population was left with a high positive activation (a case sometimes described as “winner-take-all”).

The categorizations of $F_1$ patterns made by the type of $F_2$ network just described would depend entirely on the fixed random weightings of its connections from $F_2$. Experience would have no effect on categorization. Clearly this is not the case with biological neural networks. They readily learn and adapt themselves to their particular environment. Nature has provided the mechanism for learning and adaptation in the form of the synapse, a modifiable microscopic neural juncture typically found between the outgoing axon of one neuron and the incoming dendrite of another.

MODIFICATION OF SYNAPTIC CONNECTIONS

A synapse performs two roles that are crucial for the development of structured knowledge. First, speaking metaphorically, a synapse functions as a type of valve regulating a receiving cell’s response to a particular input. In a Grossberg network, a receiving cell’s response ($R$) to each incoming excitation is defined as the product of that excitation ($E$) and the strength of the intervening synapse ($S$), that is, $R = SE$. Second, again speaking metaphorically, a synapse functions as a type of monitor that adjusts its valve in proportion to both the strength of the input signal sent to it and the level of activation in the receiving cell. In a Grossberg network, each idealized synapse is adjusted according to the following rule: for a synapse of strength $S$, an incoming excitation of strength $E$, and an activation in the receiving cell of strength $A$ ($A \geq 0$), the synapse will be changed by some fraction of the
amount $A(E - S)$, that is, $\Delta S = A(E - S)/n$. This equation is deceptively simple, involving only basic information directly available at the synaptic site. Yet it holds a key to the formation of structured knowledge. Let us explore it in some detail to see how this is possible.

Variable $A$, the activation in the receiving cell, refers to $F_2$ populations. Because of the competitive interactions across this field, only one population remains highly active after the arrival of excitation from $F_1$. This means that for only one population is $A$ of any magnitude, and only cells in that population can effect significant changes in their synapses. Learning is thus restricted to one population at a time, and different $F_1$ patterns of activation will be learned by different $F_2$ populations.

Variable $E$, the strength of the incoming excitation from a single population of $F_1$ cells, is large when its source cells are highly activated. In the illustrative example of Figure 3, the values of $E$ stemming from $F_1$ populations responding to the triangle, the circle, or the square would be large, whereas values of $E$ stemming from other populations would be very small or nil. The active $F_2$ population thus would have those of its synapses strengthened that connect to the three active $F_1$ populations and those of its synapses weakened that connect to anything else.

Variable $S$, the strength of the synapse, records the sum of previous changes caused by the other two variables. If, for example, a synapse connects to an $F_1$ population that is frequently active when that synapse’s $F_2$ population is also active, the synapse will be repeatedly strengthened. By contrast, a synapse connected to an $F_1$ population that is never active at the same time as the synapse’s $F_2$ population will have its strength eventually reduced to zero. If one thinks of an $F_2$ population as representing a category, and $F_1$ populations as representing features, then strong synapses will be found connecting the category to inputs from its central features.

The typical course of learning undergone by an $F_2$ population can be described as follows. Initially the population would have synapses of random strength concerning it to each $F_1$ population. Eventually some pattern of activation $(P)$ across $F_1$ would be multiplied by the $F_2$ population’s synapses in such a way that that population would win the competition for $F_2$ activation. The resulting large activation (variable $A$) would allow the population’s synapses to strengthen in response to strong inputs from $F_1$ (variable $E$) and to weaken in response to weak or nonexistent inputs from $F_1$. Both these changes would affect the $F_2$ population’s response to subsequent $F_1$ patterns of activation. On the one hand, that $F_2$ population would respond even more strongly and quickly than before when pattern $P$ is encountered. On the other hand, its responses to other patterns would be attenuated, because synapses for non-$P$ features would have been weakened by the effect of pattern $P$.

Synaptic changes, which in actual brain tissue are chemical or structural
in nature, persist and accumulate over time, as mentioned. The result is that through the action of their synapses, particular $F_2$ populations become closely associated with various aspects of $F_1$ populations. For each $F_2$ population, the learned pattern of strong and weak inputs stored in its synapses forms the neural equivalent of a long-term memory. And the set of all the long-term memories in $F_2$ defines the categories that the field “knows” and can recognize.

Would that categorization were that simple! Unfortunately, the neural architecture just described is inadequate for adaptation to a complex environment of patterns. Imagine, for instance, that instead of one pattern $P$ and a different pattern $Q$, there was a graded series of patterns representing many stages intermediate between $P$ and $Q$. An $F_2$ population originally responding to pattern $P$ could have its synapses shifted gradually toward pattern $Q$ by exposure to patterns that are less and less $P$-like and more and more $Q$-like. Depending on the order in which the graded patterns were presented, the category structure of such an $F_2$ network could easily become unstable, with a $P$-population becoming a $Q$-population and vice versa. One would of course desire $F_2$ categories to be somewhat plastic so as to allow for a range of variation in constituent $F_1$ features. But unstable categorization would prevent the abstraction of yet higher level cognitive constructs. Speaking figuratively, knowledge based on unstable categories would be like a house built on sand.

STABILIZING FEEDBACK IN CONCERT WITH AN ORIENTING SUBSYSTEM

At the ends of the two symbolic axons shown in Figure 3 are dark squares representing modifiable synapse-like connections. The square at the left of the figure stands for the connections on feedforward paths from $F_1$ to $F_2$. Conversely, the square at the right of the figure stands for the connections on feedback paths from $F_2$ to $F_1$. So not only does bottom-up input flow from $F_1$ to $F_2$, but also top-down feedback flows from $F_2$ to $F_1$. We might variously describe this feedback as a template, a prototype, an expectation, or a schema based on what the active $F_2$ population has learned of $F_1$ patterns of activation.

Feedback from existing categories helps to stabilize their development by simultaneously reinforcing their central features and attenuating peripheral or extraneous features. These processes take place in the interconnected feedback loops of $F_1$. Where strong $F_1$ activation and $F_2$ feedback converge, activation will increase. Activations not reinforced by $F_2$ feedback, however, will decrease because of the limit on the total activation across $F_1$. Repeated exposure to a slight variant of an established pattern does allow for gradual assimilation of the variant features. But a single exposure to a slight variant may leave its respective category unaltered.
The resonance between feedforward features and a feedback schema, what Grossberg (1976) describes as an example of adaptive resonance theory (ART), is highly stable. Such resonant feedback is, for example, used in certain electronic tuning circuits where one wants to lock on to a particular signal and hold it in spite of significant frequency or amplitude variation. Yet there are times when stability itself could be a problem. The following case is illustrative.

Consider a network’s hypothetical response to F1 populations registering the first two tones of an ascending tonic triad. An “ascending-tonic-triad” population in F2 would begin to receive excitatory input and, as it did, would send feedback to F1. This feedback would reinforce the F1 populations responding to the first two tones of the triad and prime a third population in anticipation of the third tone. Should the third tone of the triad now sound, all will be well. But music is full of surprises, and the expectations of F2 populations will not always be borne out by the F1 patterns of activation. Should quite an unexpected third tone occur, the ART network just described might suppress the new tone and encode the pattern as yet another instance of an ordinary ascending tonic triad. This is clearly in error.

In broad terms, an ART network needs to know if it has made a mistake—if it has overgeneralized or made a false prediction. That knowledge can be provided by an orienting subsystem that correlates the feedback from the upper field with the input in the lower field in order to evaluate the degree of similarity between them. Such a comparison requires that feedback and input be differentiated, even though the ART resonance between F1 and F2 requires that the two be merged. Both requirements can nonetheless be met by expanding the structure shown earlier in Figure 2. In that architecture, two layers of neurons and interneurons formed small feedback loops. A more complex architecture, three units of which are shown in Figure 5, uses three layers of neurons and interneurons to create double feedback loops. Input to F1 dominates the pattern of activation in the bottom layer, feedback from F2 dominates the pattern of activation in the top layer, and both patterns blend in the middle layer.

An orienting subsystem could make useful comparisons between activations at several different points in this three-layer design. For the simulations described below, the orienting subsystem compared activations at the left side of the middle and top layers. As shown in Figure 5, the orienting subsystem consists of a set of comparator cells (the gray cells at the far left of the figure) and another interneuron. This interneuron inhibits the comparator cells in the following manner.

Let us assume that the interneuron has a threshold (T) similar to that of the other interneurons. The magnitude of excitation from the middle layer will already be near this threshold, and if the excitation from the top layer were also near the threshold, then excitation of approximately double the threshold would reach this interneuron. As discussed earlier, it would then
inhibit the comparator cells by a factor of $T/2M \approx T/2T = 1/2$, in other words halving these cells' activations. More generally, the inhibition will be by a factor of $T/(M_m + M_t)$, where $M_m$ is the magnitude of excitation from the middle layer and $M_t$ is the magnitude of excitation from the top layer.

The effect of this interneuron on a single comparator cell depends on whether or not the cell receives the same excitation from both layers of $F_1$. Let us examine the case where the interneuron halves the comparator cell’s activations. If a comparator cell receives the same excitation from both its middle- and top-layer cells, it will reach a final activation close to that of either of those two cells (if $E_m + E_t = 2E$, then $2E(T/2M) \approx 2E/2 = E$). On the other hand, if the comparator cell receives excitation from only the middle-layer cell or the top-layer cell, it will reach a final activation that is half that of the single cell exciting it ($E(T/2M) \approx E/2$).
The overall operation of the orienting subsystem may best be understood in terms of both fields of the entire network. At times when the pattern of input to F₁ is perfectly matched by the pattern of feedback from F₂ (or when such feedback has yet to develop), the set of comparator cells will register a total activation of a magnitude close to the interneuron threshold—close, for example, to the magnitude of activation in the middle-layer cells that excite the comparator cells. But should the two patterns diverge, the magnitude of total activation in the comparator cells will begin to drop, reaching its nadir when the two patterns have no features in common. The magnitude of total comparator-cell activation, within reasonable limits, is thus a measure of the degree of match between input and feedback patterns. And if this magnitude falls below a certain threshold (what Grossberg calls the vigilance parameter), a diffuse, nonspecific signal is released in F₂ to quench whichever F₂ population is currently (and incorrectly) active (Carpenter & Grossberg, 1987). A different F₂ population then becomes maximally active and the matching process begins again. The orienting subsystem thus tempers the strong stabilizing force of the ART design by disallowing erroneous codings of lower-level patterns.

Comments About the ART Network

There are intriguing parallels between an ART network’s reactions to a mismatch and some current thinking about affect and emotion. For example, what Meyer (1956, p. 14) has called the central thesis of the psychological theory of emotions, namely “that emotion or affect is aroused whenever a tendency to respond is arrested or inhibited,” could have a physiological basis in the release of a diffuse neurochemical transmitter triggered by the mismatch of bottom-up features and a top-down schema. And Mandler’s (1982) theories about the affective significance of the search for a match between perceptual features and a cognitive schema could also be exemplified by an ART network’s operations. In Mandler’s terms, little or no affect results from simply confirming a match or a near match; in network terms, small discrepancies between F₁ input and F₂ feedback do not trigger the orienting subsystem. In Mandler’s terms, more affect results from the testing of one or more alternative schemata; in network terms, larger discrepancies between F₁ input and F₂ feedback trigger the orienting subsystem and suppress one or more F₂ populations. And in Mandler’s terms, the maximum affect results from accommodating that which cannot be assimilated into existing schemata; in network terms, very large discrepancies between an F₁ input and all existing F₂ feedback patterns may require a new F₂ population to be allocated.
The claim is sometimes made that a network with competitive interactions—mutual inhibition of the type operating at the F₂ level of an ART network—merely computes the statistical correlations of the features to which it is exposed. In the words of Rumelhart and Zipser (in Rumelhart, Hinton, & Williams, 1986, I:190), “Competitive learning is an essentially nonassociative, statistical learning scheme.” Similarly, Bharucha and Olney (1989) described perceptual learning in neuronlike networks as a process of “incrementally altering the connection strengths so as to bring the expectations generated by the network in line with the transitional probabilities of the music to which one has been exposed.” For an ART network where F₂ sent no feedback to F₁, where there was no orienting subsystem, and where F₁ did not use feedback loops, interneural inhibition, or nonlinear signal functions, such assertions would be largely true. But as Grossberg (1987) has demonstrated in analyses of competitive learning schemes, all the resonances and nonlinearities present in an ART network strongly bias its perception of its inputs and make its global reactions to its environment more categorical than merely statistical. The network actively constructs internal representations of its environment, which it then uses to interpret its environment.

Levels Three and Four: The Sequence of Interpreted Events

The top two fields shown in Figure 6 (F₃ and F₄) replicate the functions of F₁ and F₂, but at a higher level of abstraction. Whereas F₁ is a temporary store for input in the form of musical features, features that are then categorized at F₂, F₃ is a temporary store for input in the form of the temporal sequence of F₂’s categories, a sequence that is then itself categorized at F₄.

Input to F₃ comes from the categorizing cells of F₂. As F₂ categorizes different musical events, different F₂ populations become highly activated. They are connected to F₃ by simple feedforward links (in Figure 6, note the lack of symbolic synapses between these levels). These links bring excitation from the currently active F₂ population to F₃. F₃, like F₁, is a dynamic short-term memory. As excitation from newly active F₂ populations enters this short-term memory, previous F₃ activations are inhibited. These earlier activations, representing earlier musical events, will fade from short-term memory when the number of new activations (new events) exceeds the field’s transient memory span (TMS). In the simulations presented below, the TMS was usually four or five events. The pattern of activation across F₃—a pattern representing a short temporal sequence of musical events—can then be categorized in F₄ by the ART network consisting of both F₃ and F₄ in conjunction with an orienting subsystem. An active F₄ population will
thus represent a high-level categorization, perhaps, as we shall see below, a cadential schema or some other conventional sequence of musical events.

With its many nonlinearities, resonances, and feedback loops, its hundreds of cells and thousands upon thousands of interconnections, and its overall hierarchical structure, this four-level network is an object of considerable complexity. But in comparison to even a small area of the cerebral cortex, that simulated complexity shrinks to insignificance. As a model of human neurology, it is at best an idealized sketch of a few limited functions. Yet it is a testament to the fundamental brilliance of the brain’s design that even an idealized sketch of some fragment of it is capable of considerable feats of information processing. Let us turn now from the general discussion of this four-level network to an examination of how it adapted itself to a rich environment of musical features—the world of young Mozart.
Categorizing Young Mozart

Child Mozart was born into a world rich in stereotyped musical patterns. At an early age, he was able to abstract from others’ compositions, and to reproduce in his own, what I have termed (1988) the prototypes of the then-current musical schemata. It is not too farfetched to say that Mozart’s written works from age 5 to perhaps 10 catalogue the accessions to his personal collection of contemporary schemata. One hears in these pieces multiple variants of standard galant phrases arranged in ever more artful combinations.

If, like young Mozart, an ART network is also able to abstract and schematize relationships among the patterns in its environment, then when exposed to the environment of Mozart’s early works, it ought to be able to discern the very schemata that Mozart used in writing them. This is the premise behind an experiment in which a four-level network similar to the one just discussed was repeatedly exposed to Mozart’s six earliest compositions (KV1a–d, KV2, KV3).

How does one expose a network to Mozart? One could start with the raw acoustic facts of live performance. This is, after all, where Mozart started. But many, many successive neuronlike networks would be required to transform an acoustic signal into musical patterns recognizable to the classically trained musician. Furthermore, there are gaps in our knowledge of how these transformations are effected and what their end products might be. I have chosen instead to begin at a more abstract level where the patterns of input to the network’s first level are already in the form of recognizable musical features derived from Mozart’s scores. This is simplification number one.

Simplification number two reduces the temporal continuum of a composition to a discrete series of events, each event being the appearance of a new chord or a new tone in the melody. Inputs representing the particular bundle of musical features that defines an event are mathematically preprocessed to imitate the time-dependent actions of hypothetical lower-level networks. For example, suppose that a new event f has been preceded by events a, b, c, d, and e. The input for the new event will contain the features for event f as well as fading traces of the features for events a, b, c, d, and e. The network is thus exposed to Mozart in flash-card fashion, each flash card exhibiting a new event as well as fading traces of previous events.

Perhaps an actual example may clarify the type of input that the network receives. At the bottom of Figure 7 there is a simple cadence taken from Mozart’s KV1b, a tiny Allegro written when he was only 5 years old. Six events are marked on the score as a, b, c, d, e, and f. Above the music are the six corresponding states of F1. Immediately below the final state, f, are listed the 31 inputs through which excitation flows into F1. These inputs
represent rather low level musical features and could be thought of as answering basic yes-or-no questions of musical perception, questions like “Is this new tone in the melody higher or lower than the previous tone?” or “Is the bass sounding the tonic pitch at this moment?” No claim is made that these particular features constitute the core elements of music perception. They were chosen in part for their easy correlation with the musical score and for their methodologically challenging heterogeneity.
State a (the top layer of Figure 7) shows a strong trace of fa in the melody, a weaker trace of the earlier melodic la, a strong trace of a descending melodic contour, and a somewhat weak trace of the earlier fa in the bass. State b shows not only the appearance of mi in the melody, sol in the bass, and an ascending contour in the bass, but also the inhibition and consequent attenuation of the features from the previous state, state a. In like fashion, each successive state from c through f adds or changes the appropriate features while continuing to inhibit features not reinforced by new events. By state f, some old features have been repeatedly inhibited to the point where the appearance of one more new feature would cause them to disappear from F. As mentioned, this limit—here six musical events—is called the transient memory store and defines the basic capacity of F’s short-term memory.

Each new pattern of activation across F is transmitted to F₂, where individual populations compete to dominate the field. I arbitrarily allotted 25 on-center populations to F₂. The young Mozart no doubt allotted millions of his brain cells to imagining the musical patterns in his pieces, but that number would have far exceeded the capabilities of my computer, or for that matter, most any computer. There is a decided advantage to having a large number of F₂ populations. Simply put, the more F₂ populations, the more distinctions the network can draw. Conversely, the fewer F₂ populations, the more abstract must be each population’s memory of a musical event. For instance, had I given F₂ only three populations, it would have been forced to see every musical event as an instance of one of only three categories. Needless to say, I did not want to create a neuronlike network merely to learn that everything is a tonic, dominant, or subdominant. Generations of music theorists have already labored in an effort to tidy up that ground. My interests lie in how musicians deal with larger sets of less abstract patterns, patterns not constrained to fit within a too-ready systematization. Twenty-five populations seemed a set large enough to be interesting in this regard and yet still small enough to be computationally manageable on a microcomputer.

As mentioned previously, the network was taught six early works of Mozart (KV1a–d, KV2, KV3). The first four of these only came to be published in the Mozart bicentennial of 1956 and are thought by Plath (1982) and others to be his earliest compositions, displacing in that respect the better-known minuets of K1. To our ears, the pieces are musical trifles; anyone with a broad knowledge of eighteenth-century music will have heard their generic, schematic patterns time and again. But for the computer program simulating the operations of these neural fields, a program I dubbed L’ART pour l’art, everything was new—the six pieces constituted its entire musical universe. Given no previous knowledge of music (in fact, no previous knowledge of anything), it had to work hard to categorize each of the 793
separate musical events in this small repertory (each event being a new tone or chord and the decaying trace of what preceded it). Yet for neuronlike networks, just as for tyro musicians, practice makes perfect. After the pieces had been taught to the network 12 times—at varying “tempi,” in varying order, and at slightly different rates of learning—and after L’ART pour l’art had had a chance to make nearly 10,000 successful categorizations, a stable category structure emerged. This might not have been possible had the pieces contained no repetitions of material or not made ample use of stock patterns at every turn.

When learning began, F₂ was inchoate. Then it slowly organized itself, by itself, as it encountered F₁ patterns that bore family resemblances to each other. In the first stages of learning, the similarities that F₂ detected were not always ones based on chords or harmonic functions. The contour of the melody and bass, for example, exerted a strong influence on the network’s early judgments of similarity, even in cases where, as with dominant and tonic chords, the events might be harmonically disparate. This would seem to echo Dowling and Harwood’s (1986) observations about the importance of contour to the perceptions of untrained listeners and children.

Perhaps similarities in the composite contours of the melody and bass were initially important because the inputs for contour admitted only a small number of basic combinations: melody up, bass down; melody down, bass up; and so on. By contrast, the inputs for tonal features admitted thousands upon thousands of combinations. Only gradually did the network generalize the type of tonal events we describe as chords, chord progressions, or voice-leading combinations. Subtle differences in voice leading were often overlooked until F₂ memories had become considerably refined. For instance, L’ART pour l’art originally placed the deceptive cadence in the same category as the more common authentic cadence. The gross similarities between the cadences dominated their categorizations until F₂ feedback became strong enough and focused enough to point up their differences. Moreover, since difference per se is no guide to musical meaning, L’ART pour l’art had to learn which differences mattered—which features were central and which were peripheral to each category. Complicating its task was its complete lack of knowledge about the interrelationships between features. The brackets for “melody,” “bass,” and “inner voices” shown on Figure 7 may suggest that the network knew which features had a priori affinities, but this was not the case. Presented with only uninterpreted F₁ patterns of activation—peaks and valleys without accompanying labels—it took the network quite some time (as it would you or me) to notice that there was a consistently altered subset of authentic cadences—the deceptive cadences.

In observing the emerging musical categories of an early version of
L’ART pour l’art, I was troubled by how many F₁ populations seemed to be allocated to minor variants of the same events. A mere passing tone or neighboring tone could cause the network to miss the underlying similarity between two events. Of course I hear these largely two-voice pieces as implying inner voices and standard tonal schemata, so my frame of reference is broader than what was available to L’ART pour l’art. But more seemed to be involved than just a better understanding of chords and chord progressions. After all, how does one get to that level of understanding? It occurred to me that my perceptions also differed from those of the network because I hear Mozart in a metric context. If Cooper and Meyer (1960, p. 4) are correct in claiming that “in order for meter to exist, some of the pulses in a series must be accented—marked for consciousness—relative to others,” then L’ART pour l’art needed a way to mark an event for consciousness.

One could, for example, mark metric positions as features. One could add new populations to F₁, assign them to features such as “strong beat,” “weak beat,” “downbeat,” and so forth, and hope F₂ could correlate metric locations with particular tonal events. I did in fact try this approach with a PDP-type network. It was not very successful, however. In a few special cases, tonal events do seem linked to metric positions. The cadential six-four chord, for instance, is usually found on a strong beat. But in most cases, the same chord, melody, or progression can freely occur in many different metric positions. The network consequently learns to ignore populations assigned to meter.

If meter is not a temporal sequence of features, then perhaps it is a mode of attending—perhaps one could equate marking an event for consciousness with paying more attention to it. The concept requires that humans be able to synchronize their attention levels with the metric patterns of music. Assuming this is possible, perhaps through neural mechanisms of the type discussed in Gjerdingen (1989a), a strong beat then becomes a strong beat because we pay strong attention to it. To simulate this mode of attending in the context of a computer program, I gave F₁ five oscillating levels of attention to input, levels correlated by me with the music’s meter. The same musical event occurring on a strong beat consequently took a larger share of total F₁ activation than it would have on a weak beat. After the introduction of metric attention, the network showed a marked improvement in its ability to recognize similar musical events. Passing tones and other subsidiary events were interpreted more in keeping with the ordinary perceptions of musicians. After first reporting on the use of modulated attention in L’ART pour l’art (1989b), I learned of M. R. Jones’s (Jones, 1982; Jones, Boltz, & Kidd, 1986) discussions of rhythmic attention in relation to the human perception of music. I was encouraged to find support from a respected cognitive psychologist for what seemed at the time a highly speculative notion.

As previously mentioned, after 12 passes through the six pieces (L’ART
pour l'art scrupulously observed all the repeat signs) $F_2$ achieved a stable categorization of input patterns. The synaptic connections between $F_1$ and $F_2$ had modified themselves so that a given input pattern in $F_1$ would always excite the same $F_2$ population. This meant that $F_2$ populations had developed memories of $F_1$ patterns. But what were those memories? One of the great appeals of working with neuronlike networks is that one can inspect and analyze them in ways that are not presently possible with living brain tissue. One can easily locate each $F_2$ population's "synapses," accurately measure their strengths, and reliably record the population's reactions to various patterns of excitation from $F_1$. In Figure 8, I have taken this information and attempted to translate each $F_2$ population's memory directly into a type of musical notation. As with any translation, there are both gains and losses. Lost is the precise weighting of each feature and the way in which these weightings interact with a particular pattern of activation in $F_1$. But gained is the opportunity to glimpse something of the general voice-leading combination that each $F_2$ population was able to abstract from its limited musical experiences.

By equating the strength of a synaptic connection with the size of a notehead or arrow, Figure 8 provides a rough sense of the importance of a particular feature to each memory. In memory 13, for example, the large arrows show that contrary motion expanding to a final tonic consonance is a central feature of this memory. Not surprisingly, this memory was excited at the final cadence of all six pieces. By contrast, the tiny horizontal arrow over the melody of memory 24—a "cadential six-four"—shows that the pattern had weak ascending and descending melodic contours in more or less equal proportion. In the proper context of preceding events, this memory was excited by either the second or the third scale degree in the melody, as well as by the melodic scale-degree sequences $2\rightarrow3$ or $3\rightarrow2$. The fact that the network recognizes both scale degrees over a dominant bass as the same general event (preceded, of course, by the proper context) conforms with the common opinion of many music theorists concerning the cadential six-four pattern.

Memories numbers 13 and 24 are robust and focused. The patterns that excite them appeared again and again in the six pieces. Other memories are weaker and more diffuse, often incorporating elements of multiple, more exceptional patterns. Memory 8, for example, was so diffuse that as other, simpler memories strengthened, they eventually superseded it. Memory 21, on the other hand, did develop strong associations with $F_1$ patterns. But it responds to the melodic $5\rightarrow4$ descent supported by either of two distinct basses. In this regard, most of the chromatic alterations marked on Figure 8 are indications of the merged memories of several related patterns. The parenthetical sharp in the bass of memory 25, for example, shows that the network noticed the close association between what many would call a IV$^6$
Fig. 8. The “memories” that developed in the 25 F₂ populations allocated to categorizing musical events in Mozart’s early works. The largest noteheads indicate the strongest pitch traces, arrows signify traces of contour, “d” means the trace of a contrapuntal dissonance, and “π” signifies a harmonic tritone.
chord (f-a-d', the first inversion of the d-minor, supertonic triad) and V\(^\prime\) of V (f\#-a-c'-d', the first inversion of the dominant seventh of the dominant triad).

Associations of the type just mentioned relate different exemplars of the same schema. The network also indirectly calculates the similarity of an exemplar to different schemata. This information comes to light during competitive F\(_2\) interactions. Each time a new pattern of input sends a new pattern of excitation to F\(_2\), several F\(_2\) populations compete both to reach a high level of activation and to inhibit all their neighboring populations. Although only one population wins this competition, the identities of the losers provide a guide to related patterns that share musical features. In Figure 9, one can see the moment-by-moment competition of F\(_2\) populations as

![Diagram](image-url)

**Fig. 9.** Successive F\(_2\) activations (from back to front). The competition between populations eventually leads to one winning population, a few secondary populations with slight positive activations, and most others with negative activations. Those populations with positive activations share subsets of musical features.
they respond to the last event in a deceptive cadence. Population number 6 wins in the end. Its appropriateness is clear from an examination of memory 6 in Figure 8. But the several subsidiary peaks that retain slight positive activations point out other schemata in which some of the same musical features also play an important role. For example, population 13 has the memory of an authentic cadence in its synaptic connections. It responds to the cadential melody even though the 5→6 bass is beyond the range of variation for population 13’s schema. And population 9, although adapted to respond to V of ii (the secondary dominant of the supertonic), nevertheless receives considerable excitation from the presence of the vi chord (the submediant). For L’ART pour l’art, one can say that the relative levels of activation after F₂ competition indicate relative levels of musical similarity.

The ability of a network to gauge similarity in this way can assist us in investigations of historical musical styles. We today are hard pressed to hear works from the 1760s with the ears of the 1760s. Quite naturally, we hear early Mozart with ears accustomed to late Mozart, Beethoven, and so forth. L’ART pour l’art, on the other hand, only knows music from 1761 and early 1762, and so limits its judgments of similarity to patterns found in that narrow period. Permit me to describe one example of how this distinction proved useful.

As shown in Figure 10, measure 7 of the short Andante KV1a suggests, by its ungainly direct motion to the fifth on d, that a knowledge of counterpoint is not innate even among the most gifted composers. There is good evidence that KV1a is one of the earliest works by Mozart to survive. It is, of course, in the handwriting of Mozart’s father. But unlike some later pieces that his father may well have improved for publication, this piece retains the oddities that one might expect from a 5-year-old composer. The awkward counterpoint is one indication of the novice, as is the change in meter from four measures of 3/4 time to five measures of 2/4 time.

Measure 7 of this work has always bothered me, and I was surprised to see, when an earlier version of L’ART pour l’art first encountered it, that F₂ indicated a strong similarity between the last eighth note of measure 7 and the second quarter note of measure 3, as I have indicated on the figure. L’ART pour l’art, with its limited experience, had never “heard” anything like the end of measure 7 and so first tried to relate it to other, better-known patterns. On reflection, I realized that I also had never heard anything quite like this in this repertory. In equivalent C-major contexts, the low B♭ would normally be followed by an A, not a C.

If one closely examines measure 7 in the authoritative New Mozart Edition (Plath, 1982, p. 87), it is apparent that the two flats are drawn smaller than the flat at the end of measure 6. The size difference is meant to signal that the two flats in measure 7 are editorial emendations. The first one is absolutely necessary, recognizing the implied suspension of B♭ across the
barline. The second one, however, I believe to be unwarranted. Modern
readers accustomed to modern scores see a flat in one octave and presume
that without an indicated natural sign, the flat must apply in any octave.
This was not necessarily true in earlier times. The network may have been
right to associate the end of measure 7 with an implied first-inversion domi-
nant chord, for that is a much more likely pattern. A B♭ in the bass anchors
the C-major tonality and prevents the otherwise modal bass uncharacteris-
tic of galant writing. I have asked Eugene Wolf, Lawrence Archbold, and
other colleagues knowledgeable in eighteenth-century music about this
point, and they have all corroborated the network’s sense of style.

Because the F₁ patterns learned by each F₂ population all contain a trace
of the recent musical past, F₂ populations become quite sensitive to the pre-
ceding context. This is apparent in the F₂ populations that learned the vari-
ous stages of the stock galant cadence first discussed in connection with Fig-
ure 7 and reproduced in Figure 11. As the cadence unfolds, seven segments
of it, shown by the brackets beneath the music in Figure 11, elicit responses
in seven F₂ populations. Dotted lines from the brackets lead to each popula-
tion’s memory, shown in the musical form of Figure 8. One can see that
each memory does take cognizance of the preceding context. The second
memory for example, that of population 20, has the fifth degree of the scale as its strongest melodic feature but also has a strong trace of the first degree. This first degree is the salient melodic feature in the first memory, that of population 17. In turn, the third memory, that of population 23, has traces of the salient features of the second memory, and so forth.

The memory traces retained in the synapses of each $F_2$ population are not, of course, exact records of any particular musical passage. Rather they are composite abstractions derived from many related patterns. Taking the cadence in Figure 11 as a case in point, population 13 has a strong memory trace of a complete tonic triad with the third and fifth in inner voices, even though there were no inner voices in this exemplar of the cadence. But since complete tonic triads occurred in other exemplars, those inner voices became central features in the abstracted memory. If L’ART pour l’art were able to produce exemplars from its own schematic memories, it would exhibit the kind of reconstructive memory made famous in the studies of Bartlett (1932).

Fig. 11. A galant cadence as parsed by L’ART pour l’art. Numbers below the music refer to the “memories” listed in Fig. 8.
It is evident from the preceding discussion that many of the same seven cells would most likely respond in sequence to most stock examples of that type of galant cadence. Even in a piece that the network had not previously encountered, the musical unfolding of such a cadence would elicit responses in the same stereotyped sequence of F₂ populations. One can make this sequential excitation of F₂ populations the input to a still higher level neural field—F₃. Figure 12 is a computer-drawn image of all four levels of L’ART pour l’art, where F₁-F₂ and F₃-F₄ form a hierarchically connected pair of ART circuits. The figure shows L’ART pour l’art’s response to the end of the galant cadence just discussed.

F₁ depicts the state of the short-term featural memory when the cadence sounds the final tonic chord. A comparison with stage f of Figure 7 will serve to identify the individual features. F₂ shows a single population highly excited by the complex F₁ pattern of activation. As numbered in Figure 8, this would be population 13. F₃ shows this population, the one active in the present, with the highest activation; previous populations in the cadential sequence—ones numbered 4, 24, 5, and 23—have progressively lower activations. F₃ thus functions as a short-term memory more abstract and spanning more time than that of F₁. It records a temporal series of meaningful events. Finally, F₄ shows a single population that responds to the entire F₃ pattern of activation and thus can learn to recognize the entire latter part of the cadence.

Fig. 12. A computer-generated drawing of all four levels of L’ART pour l’art. The drawing depicts the network’s response to the final chord of the cadence shown in Fig. 11.
F₃ in Figure 12 also indicates a small activation for population number 2, a population not stimulated by this particular cadence. Population 2 normally responds to the ascending appoggiatura from the melodic leading tone to the tonic over a tonic bass, the penultimate event in many of Mozart’s early cadences. So the weak activation of F₃ population 2 derives not from the cadence at hand but from the schematic memory of the active F₄ population. This F₄ population developed eight significant memory traces in its synaptic connections with F₃. In ascending order of strength, they correspond to F₃ populations 18, 23, 25, 24, 5, 2, 4, and 13, with 4 and 13 having by far the strongest traces. Of course the transient memory span of F₃ is only four or five events (although feedback from F₄ can extend it an additional item or two, creating what Grossberg defines as the immediate memory span). So not all of those F₃ populations will have been active at any single movement in a given cadence. The cadence of Figures 7 and 11, for example, excited populations 17, 20, 23, 5, 24, 4, and 13. Of these, only the final five match the set of memory traces for the F₄ population, and only the last two are in the identical sequence. But the last two are the salient ones. F₄ reliably chooses the same population every time this cadence occurs because, taking all factors into account, that population’s memory has the best match with the F₃ pattern of activation.

The F₄ population just discussed could be called an “authentic cadence” detector. As one might guess, there also developed a “deceptive cadence” detector. Its memory traces record many of the same F₃ populations that occur in the authentic cadence but with the crucial population 13 (the final V-I progression) replaced with population 6 (a V-vi progression), that is, populations 17, 2, 5, 25, 4, 24, and 6.

Although F₄ was able to categorize temporal sequences of musical events in ways that often make good sense to musicians, it did not develop the acute sensitivity to the exact sequence of certain privileged events that would characterize a classical musician’s perception. For example, F₄ might view the chord sequences I-V-I-V-I and V-I-V-I-V as being nearly identical. Yet a classical musician would say that the first is harmonically closed and the second harmonically open and unresolved. The musician maintains a sensitivity to important and distinct parts even when they are embedded in similar wholes. A self-organizing neuronlike network with the ability to make such discriminations requires that F₄ populations be connected only to specific subsets of F₃ populations and that F₄ inhibitory interactions be contingent on the proximity and overlap of these subsets of F₃ populations. The result is a complex type of circuitry but one actually more realistic in its recognition of the way spatial factors would influence patterns of neural interconnection. Cohen and Grossberg (1987) describe “masking field” equations that spell out how such a network could recognize simultaneous multiple-pattern groupings. The use of a masking-field network in the F₄ level of L’ART pour l’art is a subject of current investigation.
Figure 12 presented a hierarchical picture of musical cognition. I should point out, however, that the hierarchy implicit in such a network is not a hierarchy of simple pitches. Bundles of features are recognized as constituting significant musical events, which are then recognized as occurring in schematic sequences, which may themselves be concatenated into still larger musical schemata. Above the very lowest levels of L’ART pour l’art, the units of information are not pitches but various larger units of musical knowledge, units that could be variously termed concepts, templates, prototypes, or schemata.

If one grants an affective dimension to the network’s interactions with its environment, as suggested in the discussion of Mandler’s and Meyer’s treatments of emotion, then it would be fair to say that although L’ART pour l’art does not know much about music, it nevertheless knows what it likes. It likes the familiar, the stereotypical, and that which simultaneously excites multiple schemata at different hierarchical levels. The network can be quite intolerant of a different musical style. For example, out of curiosity I exposed the network to a passage from one of Corelli’s trio sonatas. L’ART pour l’art failed to interpret most events in the Corelli, clearly recognizing only the opening and closing tonic chords. The “vigilance parameter”—the required degree of similarity between input features and learned schemata—had been set to the level used for interpreting Mozart’s early works, and this level was far too high to allow for the stylistic divergence encountered in Corelli. Setting the vigilance parameter lower subsequently allowed more of the Corelli to be interpreted as a variant of Mozart’s musical schemata. In other words, with greater tolerance of diversity, L’ART pour l’art could recognize the underlying commonalities. And only with greater tolerance would the network learn to assimilate the new musical style. It goes without saying that these principles extend beyond computer simulations into the world at large. The intolerance of diversity, whether by a machine or a society, signals the end of learning.

The self-organizing networks here described are simpler by far than the human mind, and far too simple to understand much that is important in eighteenth-century music. These networks cannot hear, they cannot play an instrument or sing, and they cannot reflect on what they are doing. They are designed in light of what is known of the cortical neuron, but so little is known of how neurons interact in large networks that the blanks in our knowledge must still be filled in with highly speculative structures and functions. Furthermore, a network like L’ART pour l’art is captive to the inputs one gives it, inputs that here provide but a caricature of musical experience. Yet for all these limitations, it does appear that self-organizing neuronlike networks show real promise. To be able to build up knowledge of music from music itself is no small feat. To be able to interpret simple pieces in
ways that make sense to musicians is a significant accomplishment. And to be able to process new events in ways that allow for an evaluation of affect is a major advance in the modeling of music’s emotional detail. The categorizations created by these networks are sophisticated in their weightings of constituent features, sensitive to preceding contexts, adaptive to environments with denser or sparser sets of features, and—because they do not fit tidily into traditional theories—suggestive of alternative ways of conceptualizing musical structure.

References


