Meter as a Mode of Attending: A Network Simulation of Attentional Rhythmicity in Music

by

Robert O. Gjerdingen

As you read this sentence, especially if you are reading it carefully enough to know that the main verb has yet to appear, you are obviously paying considerable attention to it. But should your reading lamp suddenly explode or your chair collapse, your attention would quickly turn to the pressing calamity. Reading would stop. You could not resume reading in earnest until once again you were ready to give it the lion's share of your attention.

The point of this illustration is that we have only so much attention. Consciously or unconsciously, we must apportion attention to the many concurrent cognitive tasks that require it. Yet there are often more demands for attention than we can fill. Just as power companies must ration electricity during periods of excessive demand, we must allocate attention according to priorities. Urgent demands by high-priority activities must at times be met even if they occasionally so deplete reserves that lower priority activities are forced to curtail operations.

The constraints imposed by limited attentional resources show themselves most clearly in highly demanding tasks. Playing a musical instrument, for example, requires as much or more attention than many people have to give. Novices in an ensemble may be so

intent on playing their own part that they are unable to divert
attention toward listening to others. Even skilled performers may
find sight-reading or transposition to be tasks calling for all the
attention they can muster. I remember once trying to learn to play
the organ. Negotiating the pedals required so much of my attention
that whenever I encountered a passage demanding nimble footwork,
my left hand would stop playing. Foundering in the rough seas of
Bach, only my feet and right hand found room in the attentional
lifeboat.

But what *is* "attention"? When we pay attention to
something, what are we doing? The psychological literature on
attention is large and represents decades of active research into
these questions. For the purposes of the present discussion, I might
summarize that research by describing attention as a catchword for
the many diverse aspects of human information processing that
focus, filter, or otherwise guide our perceptions of the world. For
example, at the very low level of sensory organs, physical and
neurological constraints can produce the phenomenon of attention-
like filtering. The human ear could thus be said to pay special
attention to midrange sounds if only because it is physically unable
to respond well to extremely high or low frequencies. By contrast, at
higher levels of perception and cognition, the focusing of mental
activity must be effected through more sophisticated processes.
Take the case of concert-goers enjoying a Bach fugue.
Knowledgeable listeners will smoothly shift their attention from
voice to voice as they follow the subject through its many musical
incarnations. To manage this directed flow of attention, one
presumes they must first store a mental representation of the subject
and then monitor the entire contrapuntal fabric for appearances of

the target theme. But this is certainly not a passive act constrained by some physical anatomy. The listener must actively infer, construct, search for, recall, evaluate, and modify what are purely mental entities. "Attention" thus runs the gamut from the selective transduction of sensory stimulation to something akin to thinking itself.

Monitoring the acoustic spectrum of a lively fugue for the entrance of the subject is a Herculean mental task. The listener must find ways to clean up an Aegean stable of rapidly shifting waveforms spread across a wide band of frequencies. Composers did, of course, routinely make the listener's job somewhat easier by ensuring that the entry of the subject had perceptual salience. Bach often had a voice rest before it reentered with the subject, and he generally gave a conspicuous contour, rhythm, or ornament to a subject's initial motive. Yet these tricks alone would not significantly reduce the drain on a listener's limited attentional resources. More fundamental to the strategic rationing of attention are the ways in which traditional composers and listeners have brought about a schematization of musical space and time. Instead of attending to an infinitude of possible pitches, we attend to the handful of tones in a scale. Instead of attending to an infinitude of possible moments, we attend to the handful of beats in a meter. By these means the listener's task is reduced from one of attending equally to all possible frequencies at every possible moment to one of attending differentially to tones and beats of varying importance. Thus if the fugue subject begins on a metrically strong tonic, one is more likely to detect its recurrence at some later metrically strong tonic than anywhere else.

The psychologist Mari Riess Jones has written expansively on the subject of temporally modulated attention.[1] Several of her experiments use simple musical stimuli to test whether subjects are capable of performing temporally conditioned "attentional targeting."[2] As musicians might hope, results indicate that people are quite good at synchronizing their attention with clearly presented temporal patterns. Jones's work, worthwhile reading for those interested in questions of musical rhythm and meter, is grounded on a basic observation about the interplay between organisms and their environments. She notes that a creature's world is patterned in both space and time. Those organisms capable of discerning and internalizing their world's invariant spatio-temporal patterns will find themselves at an advantage over their competitors.

For those organisms known as listeners, the musical environment is usually teeming with recurrent temporal patterns. But how do listeners synchronize themselves with these patterns? What metronome inside of us is capable of ticking to the beat of a rumba, a sarabande, or a bolero? The human body does of course have rhythms of its own. The circadian rhythm of twenty-four hours, for example, helps to determine when people feel like sleeping or rising. But not even the most hyper of hypermeters extends to a full

---

[1]See, for example, Mari Riess Jones, "Only Time Can Tell: On the Topology of Mental Space and Time," *Critical Inquiry* 7 (1981):557-76; idem, "Structural Organization of Events in Time: A Review," *Time, Mind and Behavior*, ed. F.A. Michon and J.L. Jackson (Heidelberg: Springer Verlag, 1985); and idem, "Attentional Rhythmicity in Human Perception," *Rhythm in Psychological, Linguistic, and Musical Processes*, ed. J.R. Evans and M. Clynes (Springfield, Ill.: Charles Thomas Publishers, 1986).

[2]Jones, "Attentional Rhythmicity," p. 31.

day. Within the temporal domain of ordinary meters--perhaps one to six seconds--biological rhythms appear unable to provide us with a practical timepiece.[3] So we must look to the human mind. And if it is in the mind that we fashion a musical metronome, that metronome must in all probability be constructed of interconnected neurons.

Imagine two populations of neurons. Let us call them A and B. As shown in Figure 1, population A excites population B, but population B inhibits population A. Outside input in the form of positive excitation can be sent to population A, and both populations contribute to a positive output that can be sent to still other neural populations. If a small input begins to excite population A, its internal level of activation will start to increase. This rise in activation is driven both by the external input and by the internal feedback between individual members of this neural population.[4] When the activation in population A rises above a certain threshold, excitation will begin to flow to population B and the activation of that population will also begin to increase. As it does, population B will increasingly inhibit population A, eventually forcing the activation of population A below the threshold of its output to population B. As a result, excitation will cease flowing to population B, its inhibition of population A will decrease, and the now uninhibited activation of population A will again increase. As the reader may have already surmised, the activations of these two populations will enter a regular up-and-down oscillation as the

---

[3]A point argued by the Israeli psychologist Dan Zakay in response to questions at the conference of the International Society for the Study of Time.

[4]Feedback within a population, as well as some indirect inhibition between populations, is a sigmoid function of each population's level of positive activation.
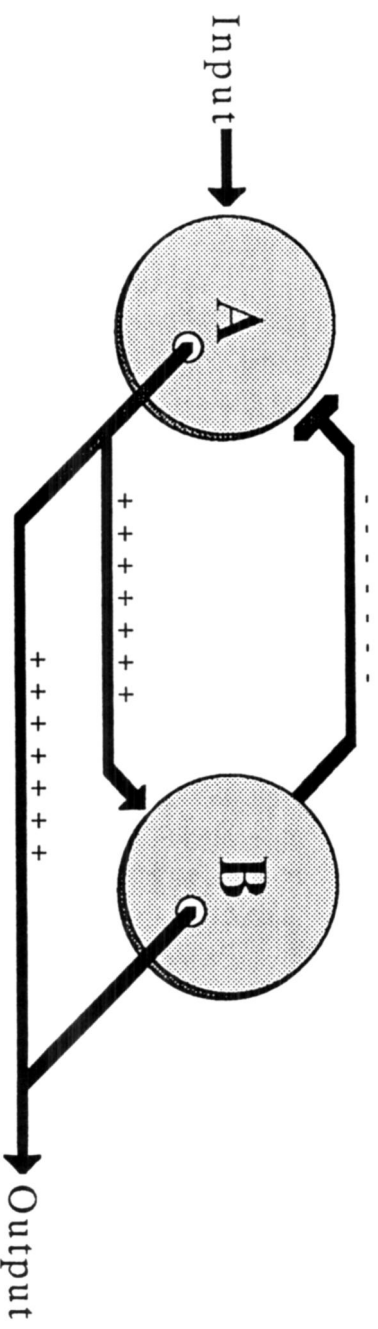
Input

A

B

+ + + + + + + + + + + + + + + + + +

- - - - - - -

Output

Figure 1. Two neural populations connected so that population A excites population B, which in turn inhibits population A.

Excitatory Output
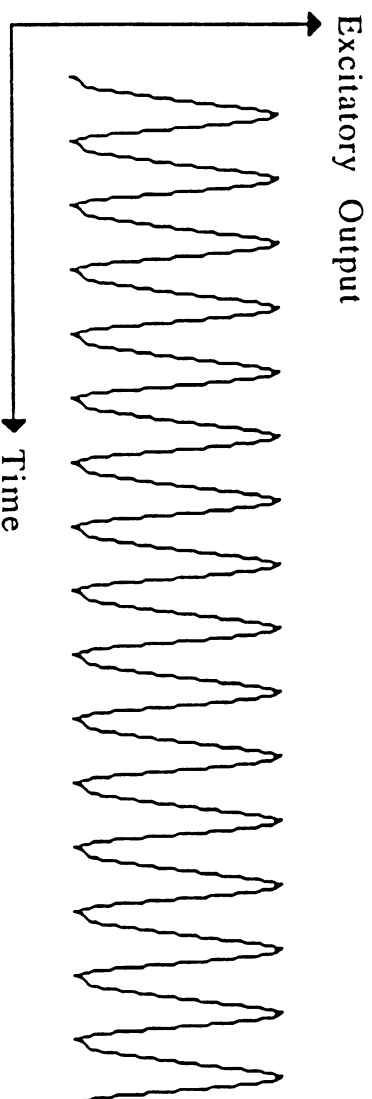
Time

Figure 2. Sinusoidal, excitatory output from the neural circuit shown in Figure 1.
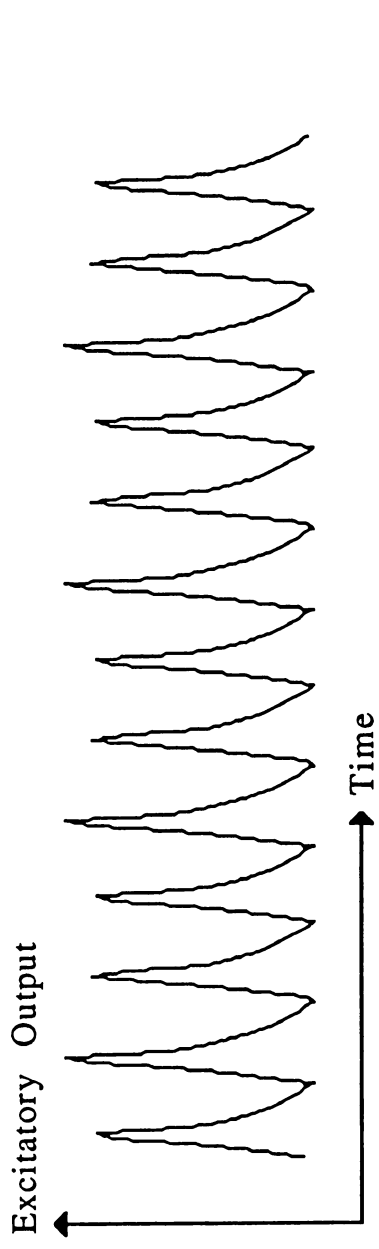
Excitatory Output

Time →

Figure 3. Excitatory output resembling "3/4 time." The longer period emerged with an increase in the flow of excitation and inhibition between the neural populations shown in Figure 1.
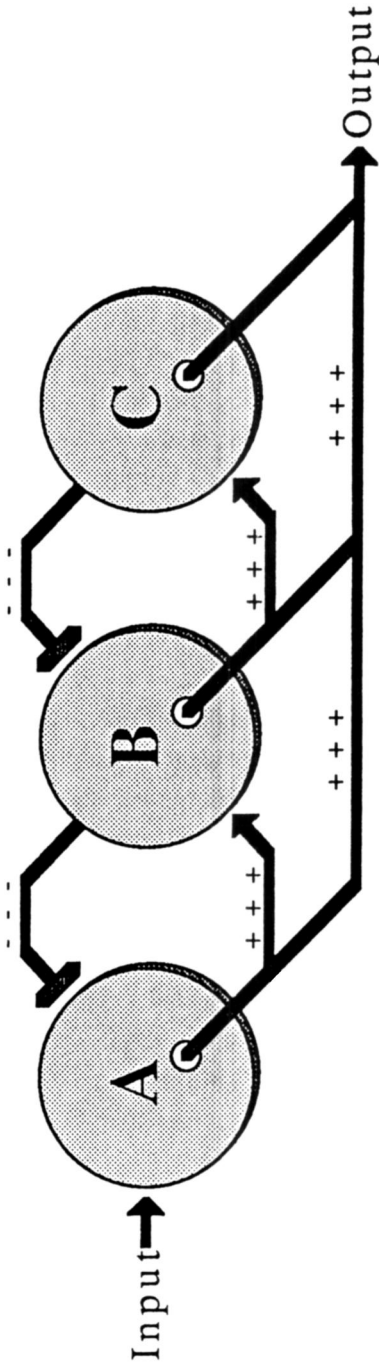
Input →

→ Output

Figure 4. Three neural populations with excitatory connections from populations A to B and B to C, and inhibitory connections from C to B and B to A.

opposing forces of excitation and inhibition ebb and flow. The two populations make up the "tick" and "tock" of a neural metronome.

Populations A and B form a complex dynamic system. Changing any of the parameters that govern its operation will likely affect the way it oscillates. In order to experiment with such alterations, I created a computer simulation of the two populations and recorded their combined output during sustained oscillation. Figure 2 shows a graph of the system's output in what I call its canonical form, that of a sinusoidal wave. The particular graph of Figure 2 is not, however, a perfect sine wave. The ascent to each crest takes longer than the descent to each trough, the ratio being about 7:5. This particular mode of neural *notes inégales* is but one of many possibilities. From the limited experiments I have done, it would appear that any ratio close to 1:1 can be obtained by minor adjustments to the system's parameters.

More substantive changes to the system can introduce more complex periodicities in its output. For instance, by doubling the rate at which excitation and inhibition flow between the populations, I obtained the output shown in Figure 3. Here the neural metronome is beating waltz time. Two different periodicites--the "quarter-notes" and the "dotted half-notes"--combine to form a composite, ternary output. Even further degrees of complexity can be achieved by adding additional neural populations. Figure 4 shows a population C added to populations A and B. With all parameters set as for the original A-B simulation shown earlier in Figures 1 and 2, the A-B-C simulation oscillates in march time, as shown in Figure 5. As before, population A excites population B, which in turns inhibits population A. Now, however, population B also excites population C, which in turn inhibits population B. Activation
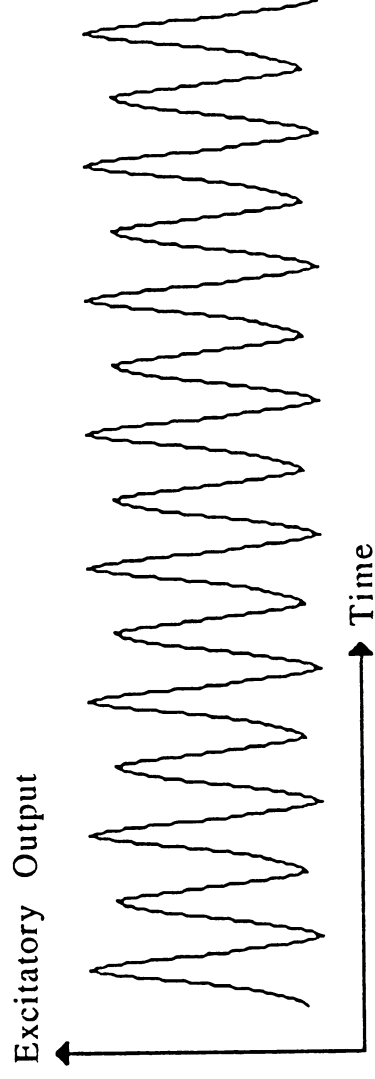
Figure 5. Excitatory, "2/4" output from the neural circuit shown in Figure 4.

reaches a maximum first at population A, then at B, and then at C, after which the cycle repeats. The greatest output occurs when both A and B have positive activations and the next greatest when both B and C have positive activations. The difference between these "strong" and "weak beats" is caused by population A having the additional excitation of its external input.

The oscillations thus far discussed were all obtained by sending each system a steady-state input. A more interesting, and perhaps more realistic, case involves sending the system a pulsating input. Figure 6 depicts the A-B-C simulation discussed above as it receives a pulsating input. At first the pulsations arrive at simulated one-second intervals. The system responds by oscillating in three-four time with each quarter-note one second in duration. As before, the "strong beats" are an emergent property of the system and not implicit in the input. Toward the right of the figure, the system is shown adjusting as the interval between pulses lengthens by about 50 percent. The smooth shift to a two-four meter shows how simply the system can accommodate a hemiola. Assuming that the pulsations were derived from the perceived rhythmicities in a piece of music, the figure suggests how easily and automatically listeners could reorient their sense of the meter.

The point of the preceding simulations has not been to suggest that anyone knows exactly how the human brain is able to synchronize itself with a musical meter. Nor is any claim made that the brain keeps the beat with these exact neural mechanisms. But the simulations do warrant two observations. The first is that neural networks oscillate as a matter of course. Anyone who simulates these complex dynamic systems on computers knows how difficult it can be to *prevent* them from oscillating. If just two groups of
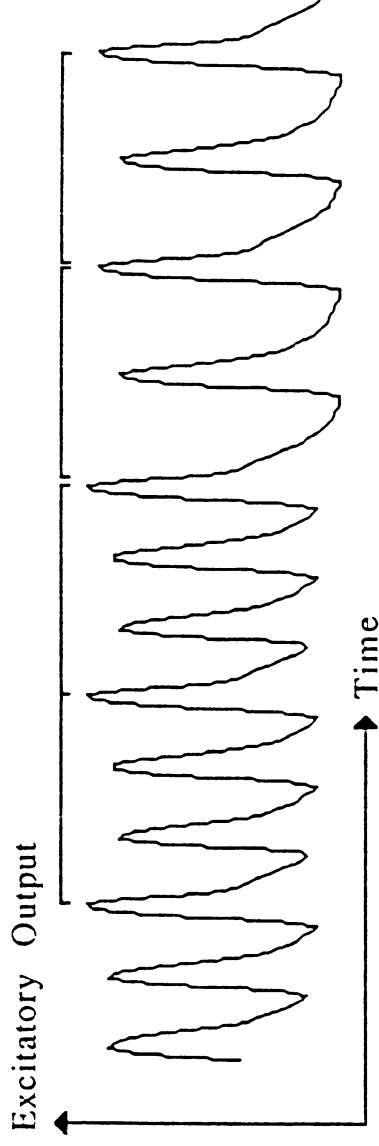
Figure 6. Excitatory output caused by pulsating input entering the neural circuit shown in Figure 4. The interval between pulses lengthens by about 50 percent toward the right of the diagram. The bracketed three- or two-pulse period is an emergent property of the circuit.

neurons can act as a type of metronome, then clearly the human brain's 10,000,000,000 neurons are capable of ticking out any meter in common use.

The second observation is that the output from a small neural metronome could be used to metrically alter the performance of a much larger neural network. Just as the tiny bias voltage from a microphone can modulate the performance of a huge amplifier, a small signal from a neural metronome could modulate the performance of a major portion of the musical mind. If such a metronomic signal were to cause regular fluctuations in a person's sensitivity to musical information, *the result would be the phenomenon of metrically modulated attention*. For the listener, musical events occurring on strong beats would then "feel" stronger than other events, even if the perceptually strong beats were objectively no louder, higher, or more strident than the weak beats. For the psychologist or music theorist observing the listener, it might well appear that the listener was consciously "paying more attention" to events occurring on strong beats, even though the phenomenon was brought about by a mostly automatic, low-level process.

To carry this speculation further, we might surmise that the ametric individual--an imagined person unable to set in motion a neural metronome--would be at a disadvantage in listening to music. While the metric individual would have heightened attention at the very moments when important musical changes are most likely to occur, the ametric individual would not. He or she would be paying equal attention to everything.

Because I am not ametric, or like to think I'm not, I find it difficult to imagine what ametric listeners would hear. The objective features of music would appear to them in the same form as they

appear to you or me. Long notes would be just as long, low notes would be just as low. Only the subjective sense of meter would be missing. One cannot, of course, run psychological experiments on hypothetically defective human beings. Nevertheless, a hypothetical ametric experience can be simulated with the aid of a computer model of a musically oriented neural network. One can compare the functioning of the network as it operates with and without metrically modulated attention.

The musically oriented neural network in question is a computer model based on Stephen Grossberg's adaptive resonance theory (ART),[5] a model I have dubbed *L'ART pour l'art*. Though its internal workings are complex and best described in more specialized literature,[6] its external behavior is quite straightforward. As shown in Figure 7, the network is composed of two levels roughly equivalent to short-term and long-term memory. As a composition is "played" to *L'ART pour l'art*, information about various musical features--scale degrees, contours, chromatic alterations, dissonances, and so forth--enters the short-term memory of the lower level. As more and more new information enters, older information is eventually forced out. The amount of music held in short-term memory thus depends on what is happening in the music. Short-

---

[5]Stephen Grossberg, "A Theory of Human Memory: Self-Organization and Performance of Sensory-Motor Codes, Maps, and Plans," *Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control* (Boston: Reidel, 1982).

[6]See Robert O. Gjerdingen, "Using Connectionist Models to Explore Complex Musical Patterns," *Computer Music Journal* (1989), in press. This issue is devoted to musical applications of neural models.

term memory will be longer for a slow, uneventful piece than for
something *molto agitato ed allegro*.

The network's upper-level attempts to recognize and
categorize the patterns of musical features that flash across the
lower level. It does this by comparing the pattern in short-term
memory with various templates or prototypes that it has learned and
stored in long-term memory. When there is a perfect match
between pattern and prototype, the network slightly adjusts its
prototype to allow for the discrepancy. And when there is not even a
rough match with any prototype, the network forms a new prototype
(if it has any memory remaining). As creator of the network, I
played no real part in these decisions. I merely set the network's
basic level of tolerance to mismatches and allotted the maximum
number of prototypes that it could form. *L'ART pour l'art*'s
knowledge of music thus derives almost entirely from the music it
has "heard." As a thoroughly self-organizing system, the network
structures itself in accordance with the structures that it can infer
from its musical environment.

In previous experiments, *L'ART pour l'art* had been taught
Mozart's six earliest pieces, the small keyboard works KV1a-d, KV2,
and KV3. For the sake of comparison, I chose to study metrically
modulated and unmodulated attention in the context of these same
pieces. The first step was to convert these pieces into a form
acceptable to the network. *L'ART pour l'art* is programmed to
perceive patterns of excitation generated by hypothetical feature
detectors. Thus real music must be converted into arrays of musical
features before it can be fed into the network. I performed this
conversion manually by going through each score and answering an
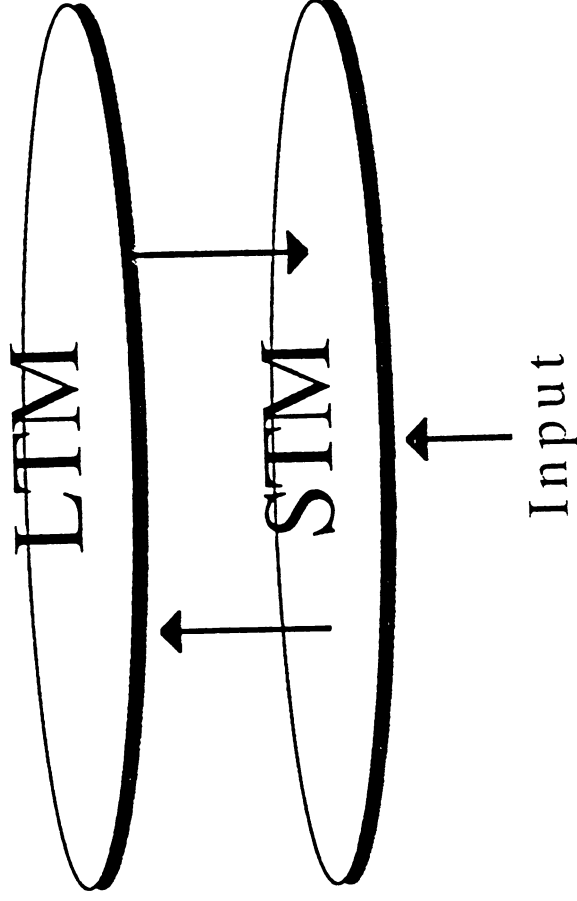array of thirty-one elementary, yes-or-no questions about each

Figure 7. Schematic diagram showing the overall design of *L'ART pour l'art*. Information about musical features enters a short-term memory (STM) where it is interpreted in the context of long-term memories (LTM).

musical event (e.g. Did the melody just ascend? Did the melody just descend? Is the melody now on the tonic pitch? Is the melody now on the supertonic pitch?, etc.). It is important to note that although I determined which feature detectors responded to which events, *L'ART pour l'art* only received an unlabeled pattern of excitation. It did not know that one excitatory signal meant something about the melody or that another meant something about the bass. All it received was an abstract, nameless pattern of excitation that might or might not resemble any pattern it had encountered before.

The next step was to scramble the network's long-term memories. By randomizing all the mathematical values that represent the associations between short-term patterns of neural activation and long-term memories, I could force the network to begin its computerized life as a perfect *tabula rasa*. And by using the same random values for both the metrically modulated and unmodulated experiments, I could force both simulations to begin from the same starting point.

The learning trials consisted of seven runs through all six pieces. Both the rate of learning--the extent to which the network could, if necessary, adjust its prototypes--and the network's tolerance to mismatches were reduced following each run. Learning was thus fastest and most literal-minded when the network began its education and then became progressively slower and more tolerant of diversity as it developed richer, stronger memories. The order in which the pieces were learned and the "tempo" at which they were presented to the network were also varied so as to prevent these arbitrary conditions from materially affecting what was learned.

Unmodulated attention is conceptually straightforward and was quite easy to simulate. The network's short-term memory was

merely set so that it received equal stimulation from the detection of a musical feature anywhere in the meter. A melodic leading tone occurring on a strong beat hence caused no greater effect than if it had occurred on the weakest part of the meter. Metrically modulated attention raises more questions. For instance, how strong is a strong beat? In absolute terms the question may have no answer. We have no unit of metrical salience, no international standard named after some great musical scholar. (Could the downbeat of Beethoven's First Symphony be 74 Sechters strong or a 6.5 on the Rameau scale?) And even in relative terms one hesitates to stipulate by how much the downbeat of a measure is stronger than the eighth-note that follows it.

In the absence of clear guidelines, one can still fall back on empirical tinkering. One can try something to see if it works, and if it does not, try something else. My course of action was to take as an arbitrary baseline the level of excitation used in the simulation of unmodulated attention. I then chose an arbitrary increment (18.75% of the baseline value) and set new levels both one and two increments above and below that baseline. As shown in Figure 8, I assigned the highest level of excitation to that produced by downbeats, the next highest to weak quarter-notes, the mid-level to weak eighth-notes, the next to the lowest to weak sixteenths, and the lowest to weak thirty-seconds and grace notes. These levels were directly tagged to the notated meter of each piece and then applied to modulate the sensitivity of the network's lower level as it received excitatory input. A musical feature occurring on a strong beat now caused a greater impact on the network's short-term memory than it would have had it occurred on any weaker beat. No attempt was made to neurally divine the appropriate meter solely from the
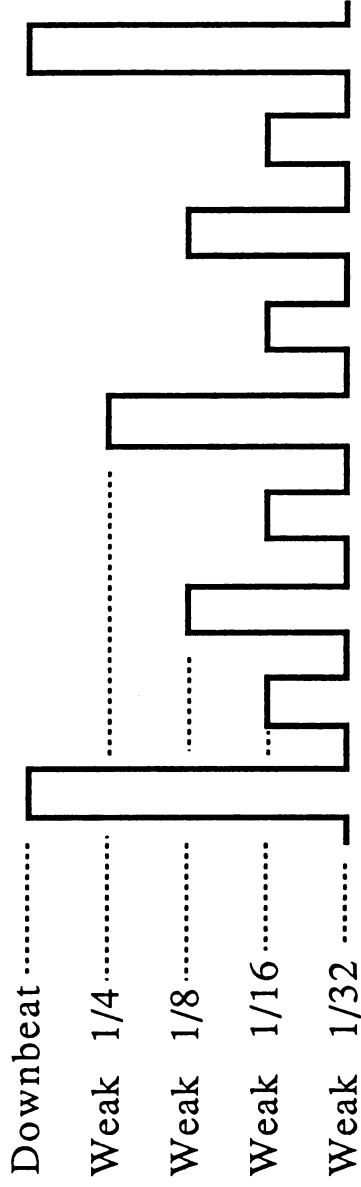
Figure 8. Plan for allocating metrically modulated attention in the case of 2/4 time. Maximum attention is given to events occurring on downbeats, less attention to events on each weaker part of the meter.

Figure 9. Comparison between metric and ametric memories. At the top is a chart of the numerical values abstracted by *L'ART pour l'art*. Below are conjectures of the musical import of those memories. Notehead size indicates salience in memory and partially translates into temporal order (large notehead = now, small notehead = before). Arrows indicate memories for contour.
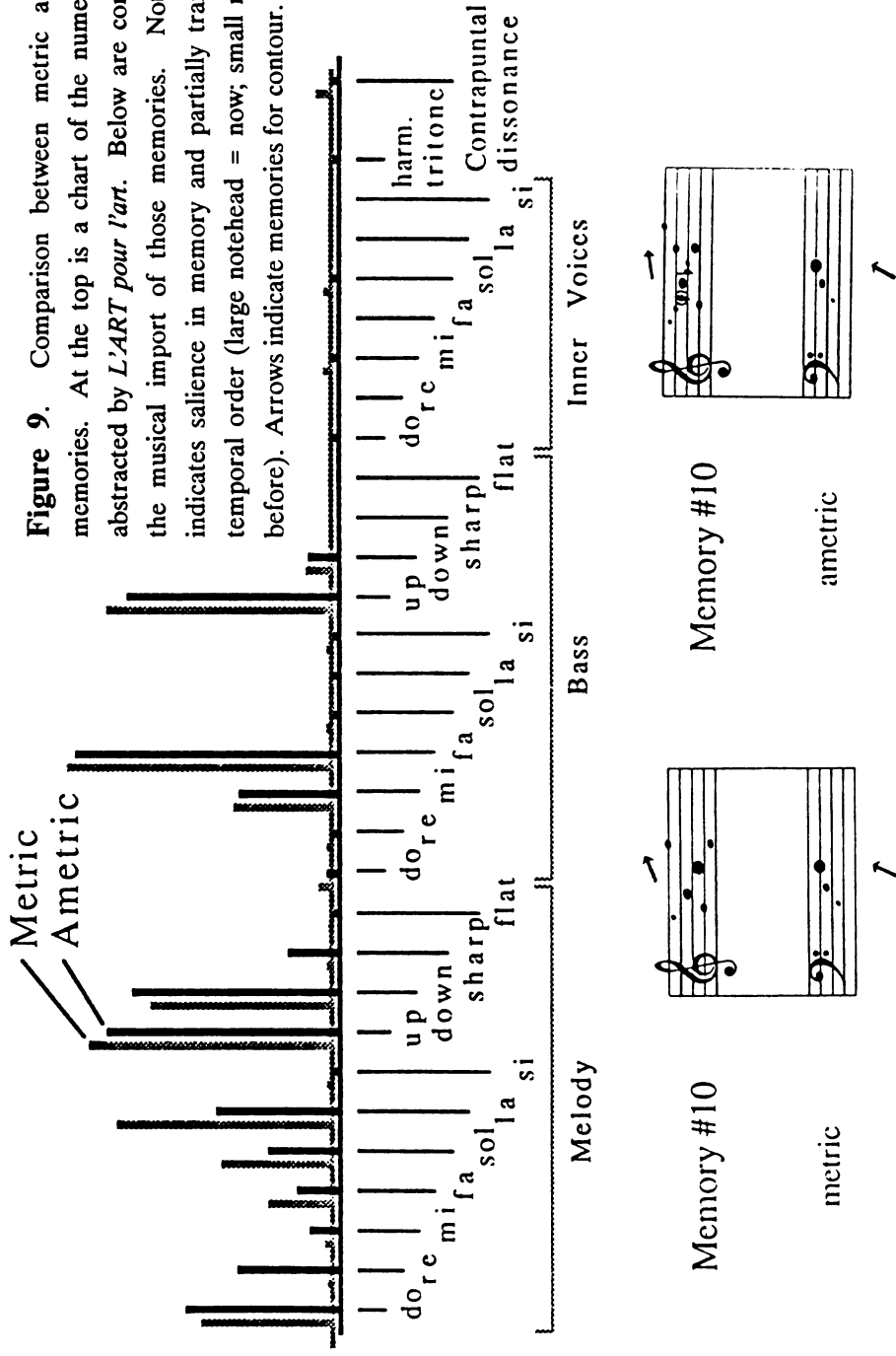
patterns of musical features. In the former case the added complexity seemed unwarranted, and in the latter case the problems involved demand a separate study.

Both the metric and ametric simulations were allowed to abstract a maximum of twenty-five prototypes from the small musical patterns in Mozart's six works. In a few cases the prototypes derived by the two experiments were nearly the same. Of course, the computer doing these simulations does not think in terms of musical concepts--everything must be a number. So from the computer's point of view, the memory of a prototype is just an array of numbers, each number standing for the strength of the association between a particular musical feature and the prototype. Figure 9 presents the similar numerical arrays representing the prototypes stored in long-term memory #10 of both simulations, first in graphic form and then in a conjectured musical form. Both versions of this prototype have abstracted the voice-leading progression from a tonic chord in first inversion to a subdominant chord in root position. The main difference between the two seems to be the richer melodic memory of the ametric simulation. The ametric simulation, more sensitive to weak passing or neighboring tones, incorporated into this prototype hints of two secondary dominants ($V^{6/5}$ of IV and vii^6 of ii) while the metric simulation associated those patterns with other prototypes.

On the sole basis of memory #10 one might assume that the ametric simulation did the better job by including more detail in its abstractions. But from a larger perspective the reverse is actually true. The metric version of memory #10 is the more focused of the two and thus less likely to be recalled at inappropriate moments. Judged by human standards, the ametric simulation consistently had

difficulty in sorting out what was and was not important. Whereas the metric simulation allotted just two prototypes to account for all the final chords in these six pieces, the ametric simulation squandered four prototypes on the same task. Whereas the metric simulation produced twenty-five distinct prototypes, the ametric simulation only ended up with twenty-four, since one of its memories was superseded by another and thus became inoperative. And whereas the metric simulation produced prototypes that always reflected an abstraction of at least two or more separate musical instances, the ametric simulation devoted one prototype of a single event occurring but once in all six works.

Mozart's early pieces are tiny by the standards of his later output. Yet counting repeats (*L'ART pour l'art* slavishly observes all repeat signs), these pieces contain nearly eight hundred separate attacks of new chords, melodic tones, or tones in the bass. To apportion each of these events into just twenty-five categories is no small task. The profligate attention to detail by the ametric simulation often left it without a suitable prototype to account for an important musical distinction. It would have to lump together musical events that a human listener would keep distinct.

Figure 10 shows measures three and four of Mozart's KV3 (1762, age 6). The numbers above the top staff show which metric and ametric memories were excited by each new eighth-note in the music. These two recognition sequences--essentially parsings of the music--begin in the same way. Memory #5 is replaced by memory #10 at the downbeat of measure three, and memory #10 is replaced on the following weak quarter. But whereas the metric simulation recognized another new event at the downbeat of measure 4, the ametric simulation did not. From the perception of the ametric

Figure 10. Measures 3 and 4 from Mozart's KV3. Numbers above the top staff show which memories were stimulated by the music. The memories diverge for numbers 20, 11, and 16, shown underneath.
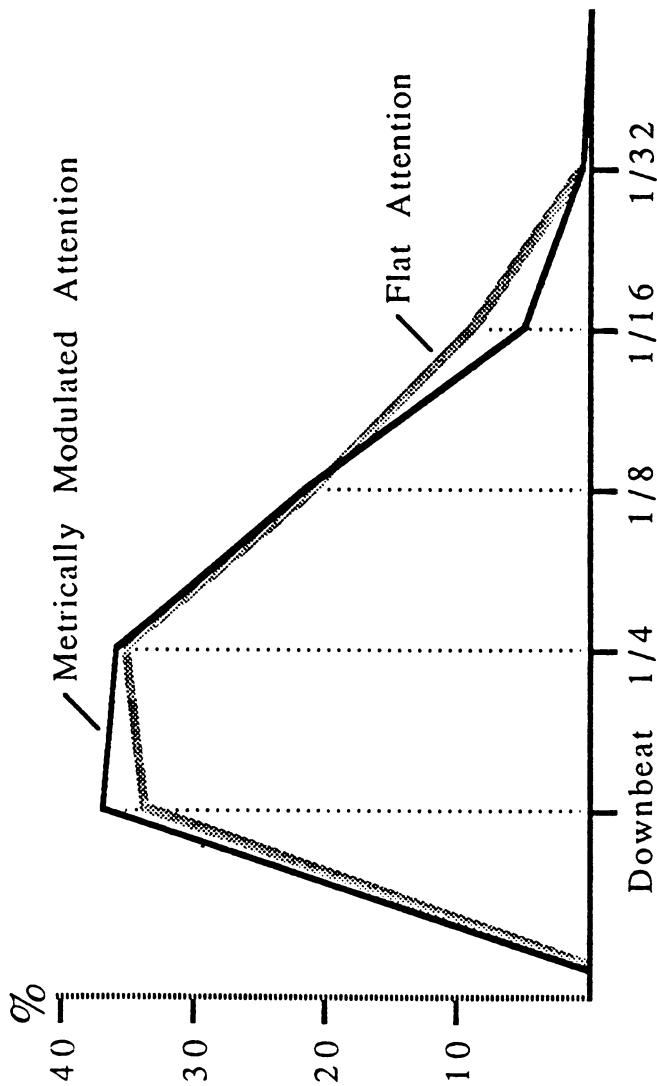
Figure 11. Comparison between simulations of metrically modulated and unmodulated, of "flat," attention. Graph shows what percentage of changes from one memory to another occurred on various parts of the meter.

simulation, the downbeat of measure 4 is an afterthought. But from the perception of the metric simulation, the downbeat of measure 4 is an important arrival at one of its two "final chord" prototypes. Clearly the metric simulation has the musically more defensible perception.

In going over the way in which each simulation parsed Mozart's compositions, I noticed several such instances where the ametric and metric simulations had different ideas about when changes in the music demanded changes in prototypes. Counting every change from one prototype to the next and noting where in the measure these changes occurred, I arrived at the statistics shown in Figure 11. In the case of metrically modulated attention, the most such changes occur on downbeats and fewer changes are to be found at each successively weaker metric location. But in the case of metrically unmodulated, or "flat," attention, the most changes occur not on downbeats but on weak quarter-notes. Furthermore, changes occur on weak sixteenth notes at almost twice the percentage in the ametric as in the metric simulation. In a curious paradox, to the simulation without a sense of meter these pieces would seem somewhat syncopated, whereas to the simulation with a sense of meter they would seem rather four-square. I suspect that anyone who knows these works would side with the metric simulation. In Mozart's earliest efforts at composition, little of significance happens anywhere but on the beat.

The fact that meter is something we can mentally create inside ourselves has tended to lead scholarly discussions of it toward notions of sophisticated cognitive processing. And perhaps in some esoteric musical repertories a measure of sophistication is indeed required to discern a meter. Yet in the vast majority of music heard

by ordinary people, meter is quite a down-to-earth affair. Could it
be that the perception of, and internal synchronization with, the
simple periodicities of basic musical features actually depends on
low-level processes? The neural metronomes discussed above are
simple "avalanche" circuits of the type found in many primitive
organisms.[7] Such circuits help the centipede reliably keep track of
which leg to move next--a task that, as the joke goes, suffers from
any higher-level processing. Could the low-level processing of meter
thus help to explain its subjectively visceral feel? And could the
overall savings in attention that result from its strategic targeting
help explain why meter is so nearly universal a feature of everyday
music? Mere simulations, even those modeling neural circuitry,
cannot provide direct answers to these questions. But simulations
do help us advance our understanding of what meter might be when
viewed as a dynamic process. And simulations can help us to see
how much our ideas of musical meaning depend on where in time
we focus our attention.

---

[7]Avalanche circuits are discussed in Stephen Grossberg, "The Adaptive Self-
organization of Serial Order in Behavior: Speech, Language, and Motor Control,"
Pattern Recognition by Humans and Machines, Vol. 1: Speech Production, ed. E.C.
Schwab and H.C. Nusbaum (New York: Academic Press, 1986).