

Cognitive Constraints on Compositional Systems

Fred Lerdahl[†]

Columbia University, New York City

In this article I explore the relationship between composing and listening. I begin with a problematic story, draw some general conclusions, introduce relevant concepts from Lerdahl and Jackendoff (1983) and related work, propose some cognitive constraints on compositional systems, discuss "pitch space", and explain why serial (or 12-tone) organizations are cognitively opaque. Most of these topics deserve fuller treatment than is given in these pages. My concern here is just to lay out a basic, if wide-ranging, argument.

I am not interested in passing judgement on the composers and compositions that are mentioned, particularly not on the remarkable work by Boulez that I use as a representative example. The thrust of my argument is psychological rather than aesthetic. But since aesthetic issues inevitably impinge on the discussion, I treat them briefly at the end.

KEY WORDS cognitive constraints, compositional systems, musical grammar, pitch space, serialism.

Hidden organization in *Le Marteau sans Maître*

Boulez's *Le Marteau sans Maître* (1954) was widely hailed as a masterpiece of post-war serialism. Yet nobody could figure out, much less hear, how the piece was serial. From hints in Boulez (1963), Koblyakov (1977) at last determined that it was indeed serial, though in an idiosyncratic way. In the interim listeners made what sense they could of the piece in ways unrelated to its construction. Nor has Koblyakov's decipherment subsequently changed how the piece is heard. Meanwhile most composers have discarded serialism, with the result that Koblyakov's contribution has caused barely a ripple of professional interest. The serial organization of *Le Marteau* would appear, 30 years later, to be irrelevant.

This story is, or should be, disturbing. There is a huge gap here between compositional system and cognized result. How can this be?

One might suppose that the impenetrability of *Le Marteau's* serial organization is due to insufficient exposure. After all, the piece was innovative; listeners must become accustomed to novel stimuli. Such has been the traditional defence of new art in the face of incomprehension. One might refine this view by pointing out that there is little repetition in *Le Marteau*. The lack of redundancy perhaps overwhelms the listener's processing capacities. Comprehensibility, then, is arguably a consequence both of the degree of conditioning to the materials and

[†] In J. Sloboda, ed., *Generative Processes in Music*, Oxford University Press, 1988.

of the number and variety of events per unit time. (Is it a coincidence that *Le Marteau* was composed in the heyday of behaviourism and information theory?)

But this explanation is inadequate. For one thing, competent listeners to *Le Marteau*, even after many hearings, still cannot even begin to hear its serial organization. For many passages they cannot even tell if wrong pitches or rhythms have been played. The piece is hard to learn by ear in a specific sense; its details have a somewhat statistical quality. Conditioning, in short, does not suffice. For another thing, *Le Marteau* does not feel structurally complex in the way, for example, that compositions by Beethoven or Schoenberg do. Vast numbers of nonredundant events fly by, but the effect is of a smooth sheen of pretty sounds. The listener's processing capacities, in short, are not overwhelmed.

This is not to deny the influence of exposure and redundancy on comprehensibility. But these factors do not go far enough; they do not address issues of specific organization. Cognitive psychology has shown in recent decades that humans structure stimuli in certain ways rather than others. Comprehension takes place when the perceiver is able to assign a precise mental representation to what is perceived. Not all stimuli, however, facilitate the formation of a mental representation. Comprehension requires a degree of ecological fit between the stimulus and the mental capabilities of the perceiver.

Experienced listeners do not find *Le Marteau* totally incomprehensible, but neither, I would argue, do they assign to it a detailed mental representation. This is why the details of the piece are hard to learn and why the piece does not in the end feel complex. The serial organization that Koblyakov found is opaque to such structuring.

Of course a musician of Boulez's calibre would not use a compositional system without drawing crucially upon his musical intuition and experience. Music-generating algorithms alone have always produced primitive outputs; not enough is known about musical composition and cognition for them to succeed. Boulez had the intellectually less ambitious goal of developing a system that could just produce a quantity of musical material having a certain consistency. He then shaped his material more or less intuitively, using both his "ear" and various unacknowledged constraints. In so doing, he listened much as another listener might. The organization deciphered by Koblyakov was just a means, and not the only one, towards an artistic end.

The degree to which *Le Marteau* is comprehensible, then, depends not on its serial organization but on what the composer added to that organization. On the other hand, the serial procedures profoundly influenced the stimulus structure, leading to a situation in which the listener cannot form a detailed mental representation of the music. The result is a piece that sounds partly patterned and partly stochastic.†

† In *Structures*, Book I (1952), Boulez flirted with "algorithmic composition", and in *Pli selon pli* (1958–61) he incorporated aleatoric elements; so *Le Marteau* lies *en route* in his dialectic of determinism and chance (see Boulez 1964). His interest in algorithmic composition continues to the extent that for a while there was a research group at IRCAM pursuing it. For six months in 1981 I was part of that group, though in my project the purpose of the computer program was not to compose but to assist composers in composing, much as a word processor assists writers (see a brief description in Lerdahl 1985). Many of the ideas expressed in this essay grew out of that experience, for which I remain deeply grateful.

Compositional and listening grammars

It will be useful to put these observations in schematic form. Let me introduce the notion of a *musical grammar*, a limited set of rules that can generate indefinitely large sets of musical events and/or their structural descriptions. Such a grammar can take the form of (a) or (b) in Figure 1. (The rectangles stand for sets of rules, the ellipses for inputs and outputs of rules.) In (a), the rule system generates the events of a passage or piece, and in the process provides some sort of specification (the "input organization"). In (b), the sequence of events is taken as given, and the role of the rule system is to assign a structural description to the sequence.

Our discussion of *Le Marteau* suggests that there are two kinds of musical grammar at work here. The first is the *compositional grammar*, consciously employed by Boulez, that generated both the events of the piece and their serial organization as discovered by Koblyakov. This grammar is an instance of (a) in Figure 1. The second kind is the *listening grammar*, more or less unconsciously employed by auditors, that generates mental representations of the music. This grammar is an instance of (b) in Figure 1; the events themselves are present when *Le Marteau* is played.

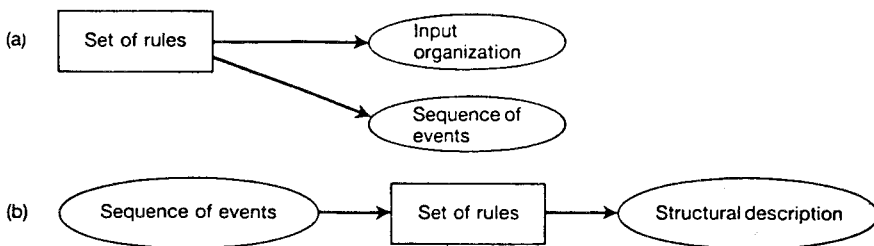


Figure 1

Figure 2 summarizes these remarks. The diagram shows that the compositional grammar generates the sequence of events and the manner in which they are specified. Only the sequence of events, however, is available as input to the listening grammar: the listener hears the acoustic signal, not its compositional specification. The listening grammar then generates the mental representation that comprises the "heard structure" of the piece. Note that Figure 2 incorporates both (a) and (b) of Figure 1.

This account is complicated by the fact that, as noted above, Boulez created *Le Marteau* not only through serial procedures but through his own inner listening. In the process he followed constraints that, while operating on the sequence of events produced by the compositional grammar, utilized principles from the listening grammar. This more elaborate picture is drawn in Figure 3.

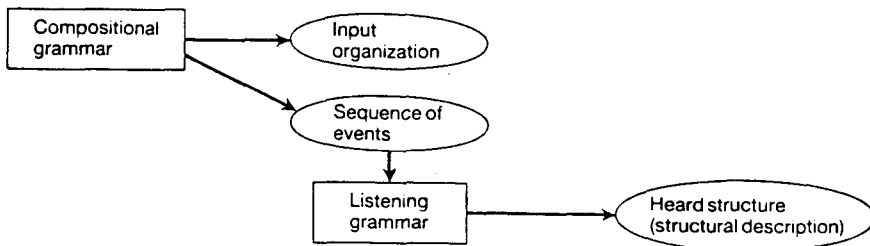


Figure 2

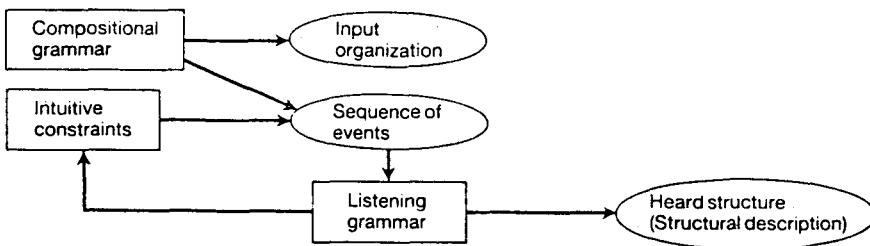


Figure 3

Observe that none of the arrows in the diagram are inputs to the “compositional grammar”. This component is inaccessible to the rest of the system. Hence it becomes quite possible for the “compositional grammar” to be unrelated to the other rules, the “listening grammar” and “intuitive constraints”. If this happens, the “input organization” will bear no relation to the “heard structure”. Here, then, lies the gap between compositional system and cognized result with which we began.

This situation exists not only for *Le Marteau* but for much of contemporary music. I could have illustrated just as well with works by Babbitt, Carter, Nono, Stockhausen, or Xenakis. This gap is a fundamental problem of contemporary music. It divorces method from intuition. Composers are faced with the unpleasant alternative of working with private codes or with no compositional grammar at all. Private codes remain idiosyncratic, competing against other private codes and creating no larger continuity – so that, for example, 30 years later the serial organization of *Le Marteau* becomes irrelevant even to other composers.

Natural and artificial compositional grammars

Where does a compositional grammar come from? The answer varies, but a few generalizations may be helpful. Let us distinguish between a “natural” and an “artificial” compositional grammar. A natural grammar arises spontaneously in a musical culture. An artificial grammar is the conscious invention of an individual or group within a culture. The two mix fruitfully in a complex and long-

lived musical culture such as that of Western tonality. A natural grammar will dominate in a culture emphasizing improvisation and encouraging active participation of the community in all the varieties of musical behaviour. An artificial grammar will tend to dominate in a culture that utilizes musical notation, that is self-conscious, and that separates musical activity into composer, performer, and listener.

The gap between compositional and listening grammars arises only when the compositional grammar is "artificial", when there is a split between production and consumption. Such a gap, incidentally, cannot arise so easily in human language. People must communicate; a member of a culture must master a linguistic grammar common to both speaking and hearing. But music has primarily an aesthetic function and need not communicate its specified structure. Hidden musical organizations can and do appear.

A natural compositional grammar depends on the listening grammar as a source. Otherwise the various musical functions could not evolve in such a spontaneous and unified fashion. An artificial compositional grammar, on the other hand, can have a variety of sources – metaphysical, numerical, historical, or whatever. It can be desirable for an artificial grammar to grow out of a natural grammar; think, for example, of the salutary role that Fux (1725) played in the history of tonality. The trouble starts only when the artificial grammar loses touch with the listening grammar.

In the Western tradition the trouble began with the exhaustion of tonality at the turn of the century. Anything became possible. Faced with chaos, composers reacted by inventing their own compositional grammars. Within an avant-garde aesthetic it became possible to believe that one's own new system was the wave of the future. Boulez's generation was the last to believe this. To a younger generation these systems have come to seem merely arbitrary. The avant-garde has withered away, and all methods and styles are available to the point of confusion.

One can react to this situation by giving up on compositional grammars and relying solely on ear and habit – on the "intuitive constraints" of Figure 3. But composing is too difficult for such a solution; there are too many possibilities. Or one can react by reverting to earlier "natural" styles, a move that condemns one to a parasitic relationship with the past. Both reactions are common these days, and both are motivated by the desire to avoid the gap between composing procedure and what is heard.

My own reaction as a composer has been less to avoid than to confront this gap. These were my initial ground rules: (1) a compositional grammar is necessary; (2) it need not be nostalgic; (3) our musical culture is too fragmented and self-conscious for a natural grammar to emerge; but (4) an artificial grammar unresponsive to musical listening is unacceptable. It was inevitable, I concluded, that early attempts at artificial grammars – say, from the 1920s to the 1950s – were defective in their relation to listening. Not enough was known about musical cognition; the basic questions had not even been framed. Beginning around 1970, however, a new perspective became possible through the simultaneous decline of the avant-garde and rise of cognitive psychology. Contemporary music had lost its way. What other foundation was there to turn to than the nature of musical understanding itself? In other words, I decided that a compositional grammar must be based on the listening grammar. Figure 4

illustrates this by adding a dashed arrow to the flow chart of Figure 3. In principle this arrow integrates the compositional grammar into the overall framework, rather than leaving it isolated from input.

But for this proposal to have substance, a great deal must be known about the listening grammar. Hence it became necessary to develop a detailed theory of musical cognition (Lerdahl and Jackendoff 1983). Such a theory, I reasoned, could provide the basis for artificial compositional grammars that could be intellectually complex yet spontaneously accessible to mental representation. The commonality of compositional and listening grammars could produce a rich yet transparent music.

Theoretical overview

This is not the place to describe the specifics of Jackendoff's and my theory. I propose only to outline some of its general features. The reader should bear in mind, however, that the theory is concrete and detailed.

Our theory models musical listening within the framework of some standard methodological idealizations. Assuming (1) an experienced listener, (2) the psychoacoustical organization of the physical signal into a "musical surface" comprising a sequence of discrete events, and (3) a final-state knowledge of the sequence, the theory predicts structural descriptions from musical surfaces by means of a set of rules that ideally corresponds to the "listening grammar" of Figures 2–4. See Figure 5; and note that the theory is "generative" in the sense of (b) in Figure 1. The predictions of the rules are testable by introspection and, in principle, by experiment.†

As a practical matter the theory focuses on Classical tonal music, but most of the rules appear to have a more general psychological basis and therefore to be applicable to other idioms as well. The extent of the universality of the rules is an empirical question. Again for practical reasons, the rules assign structural descriptions only to the hierarchical aspects of musical structure, neglecting "associational" dimensions such as motivic processes and timbral relations (but see Lerdahl 1987). By "hierarchy" is meant the strict nesting of elements or regions in relation to other elements or regions. The theory claims that, if the signal permits, the listener unconsciously infers four types of hierarchical structure from a musical surface: *grouping structure*, or the segmentation of the musical flow into units such as motives, phrases, and sections; *metrical structure*, or the pattern of periodically recurring strong and weak beats associated with the surface; *time-span reduction*, or the relative structural importance of events as heard within contextually established rhythmic units; and *prolongational reduction*, or the perceived pattern of tension and relaxation among events at various levels of structure. Both kinds of reduction are described by structural trees. The prolongational component incorporates aspects of Schenker's (1935) theory.

† Deliège (1985) has experimentally verified – and, in a few cases, improved upon – the local grouping preference rules of the theory. Todd (1985) has used the time-span reductional component of the theory to model expressive timing at cadences in performance. Palmer and Krumhansl (1987) have found empirical support for predictions made by the metrical and time-span components.

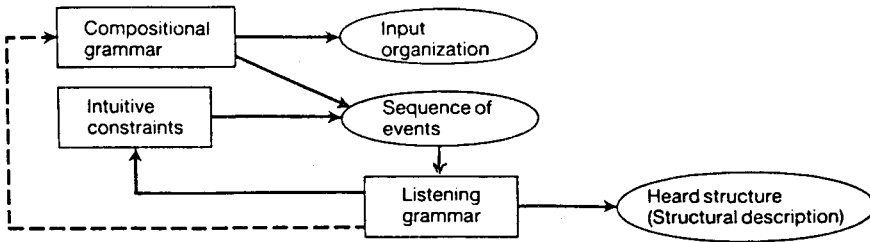


Figure 4

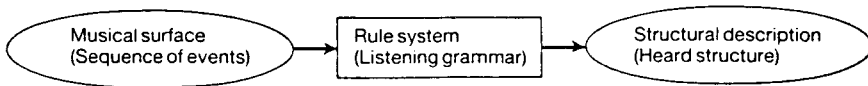


Figure 5

For the most part these four hierarchies interact as shown in Figure 6. From the grouping and metrical structures the listener forms the rhythmic units, or *time-span segmentation*, over which the dominating-subordinating relationships of time-span reduction take place; and from the time-span reduction the listener in turn projects the tensing-relaxing hierarchy of prolongational reduction. Thus the mapping from musical surface to prolongational structure is indirect. In addition, both kinds of reduction depend for their operation on *stability conditions* among pitch configurations as considered "out of time". These conditions, which also can be described hierarchically (Krumhansl 1983; Bharucha 1984b), are internal schemata induced from previously heard musical surfaces and brought to bear on the in-time event sequences described by the reductions.

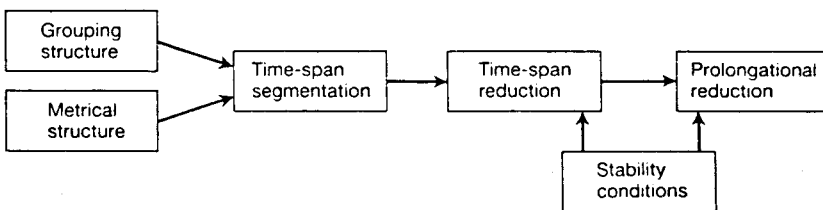


Figure 6

For each of the four organizations there is a set of *well-formedness rules* that defines the conditions for hierarchy. Well-formed structures can be modified in limited ways by *transformational rules*. These two rule types are abstract in that they only define formal possibilities. By contrast, a third rule type, *preference rules*, registers particular aspects of presented musical surfaces and selects which well-formed or transformed structures in fact apply to those surfaces. The

preference rules, which do the bulk of the work of the theory, do not generate categorically right or wrong analyses but instead predict descriptions as more or less coherent. When the preference rules reinforce one another, the analysis is stable and the passage in question is judged as stereotypical: when they conflict, the analysis is unstable and the passage is judged as ambiguous or vague.

In summary, the theory asserts that (1) the listener has at his or her disposal a number of hierarchical organizations possessing certain formal properties, and (2) the listener unconsciously attempts to assign the most stable overall description (or descriptions) by means of preferential principles that activate and interact in response to specific features of musical surfaces.

Constraints on event sequences

What does Jackendoff's and my theory have to say about closing the gap between compositional system and cognized result? Let me explore this question by proposing, in this and the following two sections, a number of psychologically plausible constraints on compositional grammars. These constraints will give substance to the dashed arrow in Figure 4. Then I will use the constraints to explain why serial organizations are not easily learnable.

We begin with a presupposition of the theory, namely that the musical surface breaks down into individual events. Certain recent musical developments (pioneered for instance by Ligeti) have tended to blur distinctions between events. Sensuously attractive though this blurring may be, it inhibits the inference of structure.

Constraint 1: The musical surface must be capable of being parsed into a sequence of discrete events.†

It is hardly fortuitous that our theory concentrates on hierarchies. Most of human cognition relies on hierarchical structuring (Miller *et al.* 1960; Simon 1962; Neisser 1967). Studies in music psychology have indicated that the absence of perceived hierarchy substantially reduces the listener's ability to learn and remember structure from musical surfaces (Deutsch 1982). It does not suffice for the input organization to be structured hierarchically; such in fact is the case for *Le Marteau* (Koblyakov 1977). It is the relationship to the listening grammar that matters.

Constraint 2: The musical surface must be available for hierarchical structuring by the listening grammar.

As suggested above, there are four ways that the listening grammar can structure event sequences hierarchically. Let us return to Figure 6 and consider what is needed at the musical surface for these organizations to come into play.

Segmentation into groups is accomplished at local levels largely by detection of *distinctive transitions* in the musical flow. A distinctive transition corresponds

† This formulation skirts the issue of simultaneous sequences of discrete events, an area that has been researched as "auditory stream segregation" by Bregman and his associates (see McAdams and Bregman 1979 for a review).

at a less fine-grained level to the principles by which events themselves are perceived as discrete rather than continuous. It is a change in some musical dimension, such as greater distance in attack points or shift in dynamics, timbre, or register, with respect to an immediate context that is relatively invariant. From this it follows that constant change will not give rise to salient distinctive transitions; nor of course will no change at all. Babbitt and Reich have something in common after all. The steady-state quality of both their musics is due to a paucity of distinctive transitions, hence of grouping boundaries.

Constraint 3: The establishment of local grouping boundaries requires the presence of salient distinctive transitions at the musical surface.

The other factor creating local groups is repetition or, more generally, *parallelism* of musical units. Repetition is an obsession of minimalism. But without distinctive transitions such groups produce only "flat" hierarchies; that is, the repetitions are not heard within larger groups to any depth of embedding. Groups of groups tend to arise rather through the reinforcing action of distinctive transitions and parallelisms, often supplemented by a third principle of *symmetry* (the approximately equal division of a larger time span). At global levels parallelism becomes the overriding grouping principle: listeners try to hear parallel passages in parallel places in the overall structure.

Constraint 4: Projection of groups, especially at larger levels, depends on symmetry and on the establishment of musical parallelisms.

Much contemporary music avoids symmetry and parallelism. Indeed, in the 1950s and 1960s literal repetition was widely regarded as aesthetically inexcusable. But symmetry and especially parallelism are basic ingredients of any complex grouping structure. The reliance on these principles in Classical tonal music and in various ethnic musics is not necessarily a sign of lack of invention but is an essential technique for the creation of deeply embedded groups.

We turn now to metrical structure. A well-formed metre is a looping pattern of equidistant beats occurring at multiple hierarchical levels. A beat at any metrical level is felt to be strong if it is also a beat at the next larger level. Depending on the musical idiom, the criterion of equidistance may be loosened, but not to the extent of abandoning all sense of periodicity. A musical idiom typically has available a limited repertory of well-formed metrical possibilities.

When hearing a musical surface, the listener seeks an optimal fit between the accents perceived directly from the signal and the repertory of available metrical structures. These *phenomenal accents* arise from local stresses such as *sforzandos*, changing bass notes and chords, contextually long events, and so forth. Insofar as possible, the onsets of phenomenal accents are aligned with relatively strong beats. In other words, the listener looks for maximal periodicity in the musical signal.

Great regularity in phenomenal accents produces a rhythmic quality of stability and squareness. In much rhythmically vital music, however, many phenomenal accents are forced to align with weak beats, creating cross-accents or syncopations. But if there is very little regularity to the phenomenal accents, the listener may not be able to infer any metrical structure. Most music is metrical,

and the immediate location of events is established largely in relation to this or that strong or weak beat. An inability to assign a metrical grid weakens the precision of location of events, resulting in a quality of suspended rhythm.

Constraint 5: The establishment of a metrical structure requires a degree of regularity in the placement of phenomenal accents.

Much contemporary music, even if notated traditionally, avoids regularity of phenomenal accents and does not give rise to a sense of metre. Complicated rhythms can cancel as well as create structure. An interesting case is Carter's practice of establishing two or more simultaneous tempos. Each such tempo corresponds to a single metrical level, or pulsation, but usually does not evoke a hierarchy of strong and weak beats. Carter contracts metrical depth in order to focus on multiple speeds. However, two or three simultaneous tempos cannot be taken in as readily as the four or five metrical levels that are commonplace in a march or waltz. Musical cognition has a bias for hierarchical organization. Simultaneous tempos instead produce independent organizations competing for attention. The difficulties in attending to more than one such organization are well known (Cherry 1953) and are ameliorated only if the vertical correspondences are coherent (Sloboda 1985), a condition that only partly obtains in this case. The situation is analogous to that of harmonically controlled versus uncontrolled polyphony.

The grouping and metrical structure come together (via well-formedness rules) to form the time-span segmentation. Details aside, it should be intuitively obvious that events are heard and thought of within such units. We normally speak of an event at small levels in terms of this or that beat, at larger levels in terms of this or that phrase or section. If the grouping or metrical structure is impoverished, then so too is the time-span segmentation.

Constraint 6: A complex time-span segmentation depends on the projection of complex grouping and metrical structures.

Various passages of *Le Marteau*, in which no metrical structure is apparent and the groups are not deeply embedded, provide a contrary illustration. Because of the consequent lack of hierarchical time-span segmentation, the sense of these passages is of a flurry of events followed by a pause, then another flurry and another pause, like beads on a string. Each flurry tends to become one complex "event" rather than an organized sequence of events. The phrase is replaced by a sonorous object extended in time. (This is one way of explaining why, despite its generally fast tempos, *Le Marteau* feels slow.)

The notion of an extended sonorous object filled with inner movement is a compelling one, especially in light of recent developments in computer music. Here lies an unexpected link between *Le Marteau* and Boulez's recent work, *Répons*. But from the perspective of musical cognition, one must ask how the inner movement of an object receives structure. We have returned to Constraint 1, the issue of musical surfaces as discrete events. If the listener can assign any rich internal structure to the sonorous object, it must be perceived as consisting of "subobjects". But if the subobjects are to be experienced as aspects of one object, they cannot be too discriminable or allow salient structuring.

The time-span segmentation must be highly structured for a complex event hierarchy to emerge – something not achieved, and perhaps not tried for, in *Le Marteau, Répons*, or indeed most contemporary music. Since “events” in music are usually pitch events, event hierarchies are normally pitch hierarchies. A unified pitch hierarchy is an instance of tonality (broadly defined), the stability conditions for which will be discussed in the next two sections. For now let us assume stability conditions and review the projection of event hierarchies from musical surfaces.

The two kinds of reduction are complementary descriptions of event hierarchies. In the time-span reduction, the listener compares the relative stability of events within each time-span segment. The less stable event in a segment is judged as subordinate to, or as an elaboration of, the more stable event. This process continues recursively from local to global levels of the time-span segmentation until an entire time-span tree is built up. The relative stability of events within a span is determined partly by rhythmic position, for example whether an event occurs on a strong beat or cadences a group. But the main determinant of domination or subordination within a span is the set of stability conditions.

Constraint 7: The projection of a time-span tree depends on a complex time-span segmentation in conjunction with a set of stability conditions.

A little thought-experiment will demonstrate the necessity of both the time-span segmentation and the stability conditions for the hearing of an event hierarchy. Imagine the pitches of *Le Marteau* plugged into the rhythms of Beethoven’s Fifth Symphony. The hierarchical time-span segmentation would largely be erased by the lack of any concomitant pitch articulation, and, in any case, the absence of stability conditions would prevent the inference of any pitch hierarchy. Or, conversely, imagine the pitches of the Fifth Symphony plugged into the rhythms of *Le Marteau*. The potentially hierarchical pitch structure would be garbled by the lack of supportive time-span segmentation (including the absence of any metrical structure or patterned grouping structure). In short, rhythm and pitch must work in concert if there is to be any time-span reduction.

The prolongational reduction derives from the time-span tree and, again, from the stability conditions. Although this is the most important of the ways in which events relate to one another, its description is rather technical and abstract, and so will be avoided here. Suffice it to say that in this component events connect hierarchically in terms of progression and continuity, creating *prolongational regions* that act in structural counterpoint to the time-span segmentation (see Lerdahl and Jackendoff 1983 for a detailed treatment).

Constraint 8: The projection of a prolongational tree depends on a corresponding time-span tree in conjunction with a set of stability conditions.

Before continuing, it would be well to emphasize how interdependent the constraints are that have been listed so far. Most of them call on their immediate predecessors. Thus a hierarchy of events (Constraint 2) requires a musical

surface that has been parsed into events (Constraint 1); the recursive segmentation of the surface into rhythmic units (Constraint 6) cannot take place without the establishment of grouping and metrical structures (Constraints 3–5); the relative structural importance of events (Constraint 7) cannot be assessed without the time-span segmentation (Constraint 6); and the representation of how events depart and return (Constraint 8) demands an evaluation of their function in the rhythmic structure (Constraint 7). The requirement of projecting event hierarchies to the listener has brought a great deal in its wake.

Constraints on underlying materials

We turn now to the stability conditions underlying the two reductions. Generally, a stability condition says, "Musical context aside, this structure is judged as more stable than that." Stability conditions interact like preference rules, sometimes reinforcing and sometimes conflicting with one another. In Classical tonality, for example, a triad in root position is more stable than a triad in first inversion, and a I chord is more stable than a V chord. These conditions reinforce one another when the judgement is between a I and a V⁶, but conflict when the judgement is between a I⁶ and a V. In actual music such conflicts are resolved by context: on the assumption that the two events come up for comparison in the same unit (time-span segment or prolongational region), the reductional preference rules decide which event is more stable on the basis of rhythmic position, linear function, and other factors.

Stability conditions are contingent on the basic materials and properties out of which the event structure of a piece is made. In Classical tonal music, for instance, the basic materials include the diatonic scale, which has certain internal and transpositional properties that can project judgements of relative stability (of which more below). Musical idioms vary in these respects. It is possible though rare for the basic materials not to be used in a way that gives rise to stability conditions: one can imagine a treatment of the diatonic scale where all scale degrees are functionally equivalent and the circle of fifths is not employed. Similarly, both Bach and Schoenberg used the well-tempered chromatic scale, but only the former carved it up so as to induce clear judgements of relative stability among pitch configurations; the latter instead sought a noncentric musical universe in which the "twelve tones . . . are related only with one another" (Schoenberg 1941). However, since learning and memory depend on hierarchical structuring (Constraint 2), and since event hierarchies depend on stability conditions (Constraints 7 and 8), it follows that stability conditions are cognitively advantageous.

What are the constraints on stability conditions? First, there must be a fixed collection of sonic elements (normally pitches; but see Lerdahl 1987). The assumption of such collections might go unremarked except that in electronic music they are fairly rare. When there is a virtual infinitude of sonic possibilities, it is hard for a composer to select and stick with a limited collection. Non-selection, however, means no syntax.

Constraint 9: Stability conditions must operate on a fixed collection of elements.

A fixed collection should not be thought of as completely rigid. First, an element of a collection need only be perceived categorically (Burns and Ward 1982); otherwise a slightly mistuned pitch would cause havoc. Second, as long as the categories remain the points of reference, there can be embellishing or sliding within or between categories (as in microtonal inflexions in Indian music or string glissandos in Western music).

The elements of a collection can be placed along a dimension to form a scale. The intervals of a scale have general size limitations; they must be large enough for adjacent elements to be easily discernible, but not so large as to use up excessive space along the continuum.

Constraint 10: Intervals between elements of a collection arranged along a scale should fall within a certain range of magnitude.

Constraints 9 and 10 are analogous to Constraint 1 (the requirement of discrete events). Both cases provide the building blocks for further organization, Constraints 9 and 10 for stability conditions and Constraint 1 for event hierarchies. A fixed collection with moderate interval sizes allows other stability conditions to emerge. In surveying some of these, let us assume that the collections are pitch collections – as opposed, say, to drums or synthesized timbres.

Most pitch collections take advantage of the 2:1 frequency ratio of the octave, so that in one respect pitches reduce to “pitch classes”. There are 88 pitches on the piano but only 12 pitch classes. Thus, in addition to giving a recurring structure to the overall collection, the octave decreases to a more manageable size the memory load for elements of the collection.

Constraint 11: A pitch collection should recur at the octave to produce pitch classes.

The octave can be divided up in innumerable ways. A common route has been to make a collection beginning with other small-ratio intervals and progressing gradually to intervals with larger ratios. From the present perspective, this route is advantageous because the resultant intervals provide a broad and graduated palette of *sensory* consonance and dissonance. Sensory consonance and dissonance can in turn form the basis for *musical* consonance and dissonance, where in a general sense consonance is equivalent to stability and dissonance to instability. Thus a seventh in Classical tonal music resolves to a sixth not just out of cultural convention but because the syntactical resolution is supported by sensory experience.

The claim is not that pitch configurations exhibiting relative sensory consonance must necessarily be more stable in musical contexts than those exhibiting relative sensory dissonance. The reverse can occur contextually, and shows up in the reductions where appropriate. Rather the claim is that the stability conditions will be relatively ineffectual unless they are supported by sensory consonance and dissonance. Stability conditions in which a seventh is supposed to be more stable than a sixth would contradict sensory experience and would not lead to event hierarchies of any depth of embedding.

Strictly speaking, these remarks hold for fundamentals with harmonic partials, such as produced by the voice and by all instruments in which the source is

periodically excited. For intervals with very small ratios, the partials of the respective fundamentals are largely reinforcing, creating the sensation of consonance. Sensory dissonance, or "roughness", arises when frequencies (whether of fundamentals or of partials) interfere within a critical bandwidth, which extends in most musical registers from a fraction of a semitone to a little less than a minor third (Plomp and Levelt 1965). The critical bandwidth is a consequence of the resolving power of the hearing mechanism in the peripheral auditory system. One could also invoke pattern-recognizing templates of the central auditory system, such as proposed in Terhardt (1974); these for present purposes lead to the same conclusion.

Constraint 12: There must be a strong psychoacoustical basis for stability conditions. For pitch collections, this entails intervals that proceed gradually from very small to comparatively large frequency ratios.

However, as is well known, small interval ratios – whether in just or Pythagorean tuning – present problems for the transposition of intervals and chords. A reading of Partch (1949) might additionally suggest that the number of pitches and intervals in such a collection can become alarmingly large. An alternative route that avoids these problems is to divide the octave into equal intervals. The collection then permits equivalence under transposition, and the pitches and intervals to be learned are limited.

Constraint 13: Division of the octave into equal parts facilitates transposition and reduces memory load.

But these advantages are offset by the fact that almost all equal divisions of the octave lead to interval ratios that are complex and that do not easily round off categorically to simple ratios. As a result the intervals give a small range of sensory consonance and dissonance, and Constraint 12 is not satisfied.

The 12-fold division of the octave (and to a lesser extent the 19-fold division; see Yasser 1932) is an exception; its intervals are sufficiently close to just intervals to be useful; that is, the deviations from just intonation are not large enough to cause serious roughness from otherwise consonant intervals. The slight loss in distinctions in sensory consonance and dissonance is compensated for by the possibility of unfettered transposition. The familiar chromatic scale has presumably survived so well because it meets the demands of both Constraints 12 and 13.

At this point two routes are available, one viewing the chromatic collection (or set) essentially as a whole, the other thinking of it as material for subsets that have various musical and psychological properties. Let us review these alternatives.

One can evoke stability conditions directly from the chromatic set through a selection and patterning of different degrees of sensory consonance and dissonance. For example, Debussy's and Bartók's harmonies often "modulate" among chord types of varying degrees of dissonance. Schoenberg was at least tacitly occupied with such a notion both early and late in his career: early, notably in the First Chamber Symphony (1906), where the basic dissonant-to-

consonant progression is from a chord built in fourths to a whole-tone chord to a triad; late, in the 12-tone works of the 1940s that include tonal features, particularly at cadential points. Hindemith's (1937) theory of "harmonic fluctuation" is based on a similar idea, generalized on a supposedly psychoacoustic foundation to embrace all possible sonorities. It is profitable to view harmonic progression in the work of Machaut and other fourteenth-century composers in such a light. Much of my own work also fits in this category.

The other route is taken by the Classical tonal system, in which the referential sonority is always the triad. Various versions of the triad alone do not project adequate differences in sensory consonance and dissonance. Stability conditions among triads are instead achieved through the use of subsets of the chromatic set, notably the diatonic. (Historically, of course, the diatonic scale precedes the chromatic, and until Wagner it was musically paramount as well. It can be illuminating, however, to view the diatonic in terms of the chromatic.)

Balzano (1980, 1982) shows through a group-theoretic analysis that only certain equal-interval sets – the 12-fold among them – lead to subsets incorporating the psychologically important criteria of intervallic "uniqueness", scale-step "coherence", and transpositional "simplicity". All three criteria have to do with location in "pitch space". The first criterion demands a unique "vector of relations" for each pitch in a subset, so that the listener can orient himself unambiguously in relation to the other pitches. For example, each scale degree of the asymmetrical diatonic subset of the chromatic set has a non-duplicating intervallic relation to the other scale degrees. In contrast, the octatonic subset (alternating half- and whole-steps), which has been widely used in this century, is not "unique", since its vector repeats four times in an octave (this is the "charm of impossibilities" described in Messiaen 1944). The second criterion ensures that the intervals of a scale proceed in an orderly manner, so that, for instance, the same distance is not traversed by one step in one part of a scale but by two or more steps in another part. The harmonic minor scale, not to mention more extreme examples, would thereby be disqualified, since the augmented second between the sixth and seventh scale degrees is equal in distance to the minor thirds that elsewhere take two steps to traverse; this presumably explains why the scale is avoided melodically in much diatonic music. The third criterion ensures orderly transposition of a subset. The diatonic subset, for example, drops one pitch and adds another as it transposes along the circle of fifths, so that distance in transposition correlates with the number of pitches shared in two diatonic scales.

Constraint 14: Assume pitch sets of n -fold equal divisions of the octave. Then subsets that satisfy uniqueness, coherence, and simplicity will facilitate location within the overall pitch space.

Balzano surprisingly demonstrates that in the 12-fold chromatic set only the pentatonic and diatonic subsets satisfy all three criteria. He advances this rather than traditional psychoacoustical explanations as the reason for the cultural ubiquity of these scales; and he recommends that the few other n -fold divisions of the octave that permit these features (such as the 20-fold set) be explored for their musical potential. However, none of the other candidates approximates just intonation. The real cognitive appeal of the chromatic set, I would argue, is

that only here do Constraints 12, 13 and 14 all converge intact. The structural features of the pentatonic and diatonic subsets would appear to be correspondingly special. (For discussion relating to this section, see Sloboda 1985; Dowling and Harwood 1986.)

Pitch space

I referred above to "pitch space". I want now to examine this notion with the purpose of extracting a few more constraints on stability conditions.

A number of cognitive psychologists, notably Longuet-Higgins (1978) and Shepard (1982), have proposed multidimensional representations for pitch relations.† The basic idea is that in a tonal framework pitches are cognized as relatively close to or far from one another, and that these distances can be represented geometrically. To take the simplest case, pitches are heard as proximate in terms of both pitch height and pitch class, so at least a two-dimensional representation is necessary. Another dimension can be added to account for fifth-relatedness; and so on. One can think of these representations as internalized maps. Just as cities are close to or far from home and can best be reached via some roads rather than others, so it is with tonal relations.

Constraint 15: Any but the most primitive stability conditions must be susceptible to multidimensional representation, where spatial distance correlates with cognitive distance.

The geometric pitch spaces proposed in the psychological literature seem promising, but I have lingering doubts. First, granting the insights provided by Balzano (1982) and Shepard (1982), it remains odd that the chromatic collection is taken as the basis for the diatonic music they seek to explain. A second and related point is that, in pursuit of geometric symmetry, these models tend to come up with more regularity than diatonic music warrants. For example Longuet-Higgins's (1978) array has a major-third axis, and Shepard's (1982) double helix is made up of two strands of whole-tone scales. But only the semitone and perfect-fifth interval cycles are relevant to diatonic music, which is in many ways asymmetrical. Other interval cycles do not begin to assume musical significance until nineteenth-century chromaticism.

These questions lead me to suggest a complementary pitch space that has a reductional rather than a geometric format. This space is musically obvious yet has explanatory power. I take the lead from Deutsch and Feroe (1981), who speak of different pitch "alphabets" – chiefly the chromatic, diatonic, and triadic – that the tonal listener "highly overlearns" from previously heard musical surfaces. This in itself is an important point, since without repeated exposure to such regularities, listeners could not spontaneously organize new input in terms of them.

† Actually, there is a long tradition in music theory for such models, dating back to Weber (1824) and his eighteenth-century predecessors; see Werts (1983). Schoenberg's (1954) chart of regions, which has been cited in the psychological literature, can be traced through Riemann back to Weber.

Constraint 16: Levels of pitch space must be sufficiently available from musical surfaces to be internalized.

Deutsch and Feroe show that memory load is greatly reduced if tonal melodies are structured by means of their alphabets. As their formalism suggests, such alphabets are hierarchically related as in Figure 7: the octave level *a* elaborates into the perfect-fifth level *b*, which elaborates into the triadic level *c*, which elaborates into the diatonic level *d*, which elaborates into the chromatic level *e*. In addition, pitches at higher levels in the hierarchy are more stable than pitches that do not appear until successively lower levels. This correlates with Krumhansl's (1979) experimental results. So Figure 7 includes a partial representation of stability conditions.

| | | | | | | | | | | | | | |
|------------------|---|----------------|---|----------------|---|---|----------------|---|----------------|---|----------------|---|---|
| Level <i>a</i> : | C | | | | | | | | | | | C | |
| Level <i>b</i> : | C | | | | | | G | | | | | C | |
| Level <i>c</i> : | C | | | E | | | G | | | | | C | |
| Level <i>d</i> : | C | D | E | F | G | A | B | C | | | | | |
| Level <i>e</i> : | C | D _b | D | E _b | E | F | F ₂ | G | A _b | A | B _b | B | C |

Figure 7

This pitch space circumvents the reservations mentioned above. First, the overall orientation is not chromatic. Non-well-tempered distinctions can be incorporated merely by expanding the content of level *e*, leaving the other levels untouched. Second, level *e* expresses an equal interval cycle and other unwanted equal divisions of the octave are absent.

As chords and diatonic collections change, so too must the hierarchy in Figure 7. For example, in the framework of a IV chord in C major, levels *a*, *b* and *c* would shift respectively to F-F, F-C-F, and F-A-C-F. If the diatonic collection changes to that of F major, level *d* would also shift, with B flat replacing B. The structure of this shifting pitch space demands further representation, perhaps geometric in format, but we will not pursue this here.

Each level in Figure 7 can be thought of as a scale with stepwise motion occurring between adjacent members. An octave is a step away at level *a*, as is a fifth at level *b*; an arpeggiation proceeds by step at level *c*; diatonic and chromatic progressions are by step at levels *d* and *e*. A skip is a progression between two elements that requires more than one step for traversal. This step/skip distinction is important for establishing cognitive distance, and was implicit in Constraint 14 above.

Figure 7 can further be seen as expressing preferred melodic routes via the Gestalt laws of proximity and good continuation. A preferred route is one that is "complete", where completeness is defined as a stepwise progression at any level that begins and ends at elements represented at the next higher level. Thus a complete chromatic progression must begin and end on diatonic pitches, and so on. Anything else – a chromatic appoggiatura, for instance – is felt as

“incomplete”. Generally, incompleteness is more acceptable at the start than at the end of a progression; unstable elements must be subsequently “anchored” (Bharucha 1984a).

Constraint 17: A reductionally organized pitch space is needed to express the steps and skips by which cognitive distance is measured and to express degrees of melodic completeness.

The notion of completeness ties in significantly with current music theories that are otherwise dissimilar. Meyer’s (1973) and Narmour’s (1977) implication-realization theory depends exactly on the space in Figure 7: a stepwise motion at any level implies the next step at that level; such an implication can be realised or not, as the case may be. And Schenker’s (1935) central conception of the *Zug*, a diatonic progression bounded by pitches of the prolonged harmony, falls out as a special case of completeness at levels *c* and *d* in Figure 7. Other Schenkerian concepts (coupling, arpeggiation, and initial ascent, the *Urfinie*, the *Bassbrechung*) can be illuminated in related fashion. Completeness appears to be the basic voice-leading principle in Classical tonal music.

Two general psychological points may be inserted here. First, even if – as at the fifth, triadic, and diatonic levels – intervals along a level are psychoacoustically unequal, they are cognitively equal because they are steps at that level. Second, because each of these levels constitutes a category in terms of which the next level is understood, Miller’s (1956) restrictions on memory load (“seven plus or minus two”) are easily met, something not accomplished by the single category of the chromatic set.

Finally, Figure 7 reflects certain psychoacoustic facts. In general, a descent from level *a* to level *e* brings increasing sensory dissonance. Level *a* gives the “virtual pitch” (or root) of the collections in levels *b* and *c* (Terhardt 1978). Level *b* serves as the harmonic norm for medieval music and various ethnic musics, level *c* for Classical tonal music. Level *d* provides the melodic basis for many musical idioms, with level *e* offering inflectional possibilities. Thus between levels *c* and *d* there appears a conceptual line separating harmony and melody. This differentiation may in part stem from the critical bandwidth, since for most registers steps at level *c* fall outside the bandwidth and steps at level *d* fall inside.

Figure 7 embodies all of the constraints on stability conditions enumerated in the last two sections: fixed collections with appropriate interval sizes (Constraints 9 and 10); octave equivalence at level *a* (Constraint 11); increasing sensory dissonance from level *a* to level *d* (Constraint 12); equal division of the octave at level *e* (Constraint 13); uniqueness, coherence, and simplicity at level *d* (Constraint 14 – and, incidentally, uniqueness and coherence at level *c*); multidimensional representation expressing cognitive distances (Constraint 15); levels of pitch space that are easily induced from a wide variety of musical surfaces (Constraint 16); and steps, skips, and degrees of melodic completeness at multiple reductional levels (Constraint 17). The cultural persistence of the many musical idioms relating to Figure 7 may be due to this convergence.

Some of these constraints seem to me binding, others optional. Constraints 9–12 are essential for the very existence of stability conditions. Constraints 13–17, on the other hand, can be variously jettisoned. The resulting stability conditions may be weaker, but they can still lead to hierarchically rich music. For

example, South Indian music approximates just intonation and does not modulate, thereby ignoring Constraint 13 and part of Constraint 14; Debussy, Bartók, and others have developed consonance-dissonance patterns directly from the total chromatic, thereby ignoring Constraints 14–17. It will be fascinating to discover how the new sound materials of computer music will be able to meet these constraints. In all probability the new material will bring additional requirements into play. Meanwhile there remains ample leeway within the chromatic collection and within the constraints listed above for music as yet unimagined.

Cognitive opacity of serialism

Now we are in a position to see why serial organizations are inaccessible to mental representation. Rather than explore the subject through the idiosyncratic serialism of *Le Marteau*, I will refer to elementary aspects of the Classical 12-tone system, which has been more widely influential (Schoenberg 1941). I will give three causes for serialism's cognitive opacity.†

Before proceeding, I must emphasize that the issue is not whether serial pieces are good or bad. As with tonal music, some serial pieces are good and most are bad. Nor am I claiming that listeners infer no structure at all from musical surfaces composed with serial techniques. What listeners in fact infer from such surfaces is an interesting question, one that deserves theory and experiment in its own right. But this is not the issue here. The issue is why competent listeners do not hear tone rows when they hear serial pieces.

The first reason is that serialism is a permutational rather than elaborational system. Pitch relations in virtually all "natural" compositional grammars are elaborational. Take for example the pitch sequence B-E-F sharp. It is easy to imagine any number of melodic embellishments internal to the sequence. This typically happens when children sing and performers improvise. Like the syntax of a sentence (Chomsky 1957), musical elaborations can continue recursively to an indefinite level of complexity; for example, sonata form is an expansion of the Classical phrase (Schenker 1935; Rosen 1972; Lerdahl and Jackendoff 1983). This feature enables pitch relations to be described hierarchically by a tree notation.

Serialism instead depends on specific orderings of the elements of a set. Distinctions arise from permutations of the elements; for example, the inversion of a 12-tone row has the same 12 pitch classes but in a different order. The order position of the elements is therefore essential to the identity of individual set forms. From this it follows that internal elaboration of any element by other elements will undermine a set form. Consider the row of Schoenberg's Violin Concerto (1936), shown in Figure 8 with order numbers given above. The sequence B-E-F sharp appears at order numbers 3–5. If one wanted to elaborate E, say, with G or with B flat-A, creating the sequences B-E-G-E-F sharp or B-E-B flat-A-E-F sharp, the integrity of the row would be destroyed, since G, B flat,

† It must be said that, if only from the evidence in Schoenberg's 12-tone sketches (Hyde 1980), Schoenberg's (1941) account of the 12-tone system is partly misleading. Plainly he used the system not just to generate rows but to create certain systematic but non-serial relationships among subsets. This, however, does not affect my argument, which concerns the cognitive opacity of serial structures (tone rows) as such.

and A have order positions elsewhere in the row. A tone row is not an elaborational structure. (The same point could be made with reference not to elements but to the intervals between elements.)

The little research that has been directed towards serialism (as in Francés 1958; Dowling 1972; Deutsch 1982; Bruner 1984) supports the contention that permutational structures are hard to learn and remember. Since other human activities are not organized in such a fashion, it is hardly surprising that the issue has in general been ignored by psychologists. The situation reminds me, though, of Chomsky's (1965) observation that many logically possible grammatical constructions never occur, such as forming interrogatives from declaratives by word reversal or by exchange of odd and even words. There is unfortunately no one on whom to test these linguistic counter-examples. Music offers a unique opportunity in this regard. Surely children can be found who have been raised on a steady diet of serial music. Do they identify tone rows? A negative answer will provide strong counter-evidence concerning the structure of musical cognition, and may suggest inherent limitations on cognitive organization in other domains as well.

Ironically, Schoenberg was much preoccupied with the issue of comprehensibility. I suspect this is one reason why in his 12-tone phase he adopted Classical motivic, phrasal, and formal structures. As a result, his serial music satisfies the rhythmic Constraints 3–6. But the permutational basis of his pitch organization assures a gap between the compositional and listening grammars.

It is of course possible to organize the combinations and sequences of individual rows on a hierarchical rather than permutational basis. Boulez, Babbitt, and others have done just that. But these higher-level hierarchies are extremely difficult to cognize in a hierarchical fashion because their underlying basis, the row, remains non-tree-like. There are further reasons as well, which brings us to the other causes of serialism's cognitive opacity.

The second cause is the avoidance of sensory consonance and dissonance as an input to the system (Constraint 12). This avoidance in turn stems from the historical collapse of the stability conditions underlying Classical tonal pieces. The pre-serial atonal period had established an aesthetic in which hierarchical pitch relations gave way to purely contextual, associative ones. The 12-tone system reinforced this trend through constant reshuffling of the total chromatic and through constant repetition of the intervallic patterns of the row. Concerned about the natural basis of his new aesthetic, Schoenberg (1911, 1941) invoked the doctrine of the "emancipation of the dissonance": consonance and dissonance are not opposites but are on a continuum determined by the overtone series; dissonances are harder to comprehend than consonances; but acquaintance with the more remote overtones has made dissonant intervals just as comprehensible as consonant intervals; so now intervals may be treated as equally consonant (or dissonant). Babbitt (1965) goes a step further, debunking the overtone series as a basis for pitch systems and claiming that consonance-dissonance distinctions among intervals are entirely contextual in origin.

Neither Schoenberg nor Babbitt distinguishes between sensory and musical consonance and dissonance. Though the details of pitch perception are not fully resolved, there is no doubt about the objective existence of sensory consonance and dissonance. For intervals between pitches with harmonic partials, the degree of sensory dissonance bears an important relationship to the overtone

| | | | | | | | | | | | | |
|---------------|---|----------------|----------------|---|---|----------------|---|----------------|---|----------------|----|----|
| Order number: | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 12-tone set: | A | B _b | E _b | B | E | F ₂ | C | C ₂ | G | A _b | D | F |

Figure 8

series, by virtue of the degree of interference among partials within the critical bandwidth. Thus far Schoenberg, at least, is right. What one does with this musically is another matter. Classical tonality founds musical consonance and dissonance largely on sensory consonance and dissonance (partial exception must be made for the perfect fourth, the minor triad, and so on). Serialism, on the other hand, treats intervallic combinations in terms of the row and in effect defines musical consonance and dissonance out of existence (with the glaring exception of the octave: pitch classes are needed). But this strategy, while perfectly logical, does not neutralize sensory consonance and dissonance. The sensory dissonance of a seventh remains greater than that of a sixth, regardless of the musical purposes to which these intervals are put.

There is an important psychological consequence: by ignoring sensory distinctions, serialism creates musical contexts that are not apprehended hierarchically. One pitch or hexachord may be associated with another pitch or hexachord, but these relationships are not easily heard in a dominating-subordinating manner. The requisite intuitions of stability and instability are missing. And because hierarchies are not inferred, tone rows and their combinations are difficult to comprehend (Constraint 2).

The third cause is that serialism does not induce a pitch space where spatial distance correlates with cognitive distance (Constraint 15). Assume for simplicity that serial pitch space looks like the "combinational" space shown in Figure 9 for Schoenberg's Violin Concerto. Here the first six pitch classes of the inverted set, transposed up five semitones, are identical to the second six pitch classes of the prime set, though in a different order. For a variety of reasons such a space was useful to Schoenberg (see Schoenberg 1941). A complete 12-tone pitch space would include transpositions of this and other relationships, just as the tonal space of Figure 7 would have to be enlarged for other chords and regions.

| | | | | | | | | | | | | |
|-----|---|----------------|----------------|---|---|----------------|---|----------------|---|----------------|---|----------------|
| P0: | A | B _b | E _b | B | E | F ₂ | C | C ₂ | G | A _b | D | F |
| I5: | D | C ₂ | A _b | C | G | F | B | B _b | E | E _b | A | F ₂ |

Figure 9

One might suppose that a "step" in such highly chromatic music is a semitone. But this is not reflected at all in the pitch space of Figure 9. Or, more in the spirit of the row, one might say that a "step" occurs from any pitch class to any horizontally adjacent pitch class. But in this case step distance varies wildly from one adjacency to the next, and there is no correlation with psychoacoustic (log frequency) distance. In short, serialism does not incorporate any psychologically coherent notion of step and skip. The listener consequently has difficulty locating a pitch as close to or far from another pitch.

Further causes of serialism's cognitive opacity could be adduced, such as its failure to provide consistent exposure from piece to piece to a limited number of "alphabets" (Constraint 16). But the main points have been addressed. We turn now to some aesthetic issues that have been lurking behind this entire discussion.

Comprehensibility and value

There is no obvious relationship between the comprehensibility of a piece and its value. Many masterpieces are esoteric, while most ephemeral music is all too comprehensible. On the other hand, if a piece cannot be understood, how can it be good? Most would agree that comprehensibility is a necessary if not sufficient condition for value.

Care must be taken with this formulation in three respects. First, comprehension presupposes listening competence for the music in question. This competence varies with ability and especially with exposure, but is not less real for that. Second, comprehension pertains to the listening grammar rather than to the compositional grammar. A serial piece may be understood in non-serial ways. Third, we are talking about intuitive rather than analytic comprehension. Along the lines of Fodor (1983), the mind's music module must be able spontaneously to form mental representations of musical structure from musical surfaces. This is quite different from using the all-purpose reasoning faculty to figure out the structure of a piece.

With these provisos in mind, I think the above formulation stands. Appreciation depends on cognition. I now want to go an aesthetic step further.

Aesthetic Claim 1: The best music utilizes the full potential of our cognitive resources.

This seemingly innocuous statement carries weight because a great deal is becoming known about how musical cognition works. I have outlined aspects of this understanding in the constraints presented above. Following them will not guarantee quality. I maintain only that following them will lead to cognitively transparent musical surfaces, and that this is in itself a positive value; and, conversely, that not following them will lead in varying degrees to cognitively opaque surfaces, and that this is in itself a negative value.

This stance can be refined through the notion of musical "complexity", which is to be contrasted with musical "complicatedness" (Lerdahl 1985). A musical surface is complicated if it has numerous non-redundant events per unit time. Complexity refers not to musical surfaces but to the richness of the structures inferred from surfaces and to the richness of their (unconscious) derivation by the listener. For example, a grouping structure is complex if it is deeply embedded and reveals structural patterns within and across levels. The derivation of such a structure is complex if all the grouping preference rules come into play and if they conflict with one another to a certain degree. (Total reinforcement of the rules would produce a stereotypical grouping structure; total conflict would create intolerable ambiguity.) I take complicatedness to be a neutral value and

complexity to be a positive one. Many musical surfaces meet the various constraints, but only those that lead to complexity employ “the full potential of our cognitive resources”.

All sorts of music satisfy these criteria – for example, Indian raga, Japanese koto, jazz, and most Western art music. Balinese gamelan falls short with respect to its primitive pitch space. Rock music fails on grounds of insufficient complexity. Much contemporary music pursues complicatedness as compensation for a lack of complexity. In short, these criteria allow for infinite variety, but only along certain lines.

I find this conclusion both exciting and – initially, at least – alarming. It is exciting because psychology really does have something substantive to say about how music might be; here is the foundation I was seeking. It is alarming because the constraints are tighter than I had bargained for. Like the old avant-gardists, I dream of the breath of other planets. Yet my argument has led from pitch hierarchies (Constraints 7–8) to an approximation of pure intervals (Constraint 12), to diatonic scales and the circle of fifths (Constraint 14), and to a pitch space that prominently includes triads (Constraint 17).

However, the constraints do not prescribe outworn styles. Rather they provide a prototype (Rosch 1975). Let me first give an uncontroversial rhythmic example. A musical surface in which the note values are multiples of 2 is intrinsically more stable and easier to cognize than one in which the note values are multiples of 7 and 11. This does not mean that the latter surface is somehow impermissible. It instead amounts to the observation that, because of the resultant ease in forming a metrical structure, note values that are multiples of smaller prime numbers are easier to process and remember, and therefore that multiples of 2 (or 3) inevitably remain a cognitive reference point for more complicated rhythms. I claim that a similar situation holds for pitch: the structure in Figure 7 (with its hierarchically organized octave, fifth, triad, and diatonic scale) remains a reference point for other kinds of pitch organization, not because of its cultural ubiquity but because it incorporates all of the constraints developed above. I take this as a given, with or against which a composer can play creatively. Of course one may opt for a less constrained pitch space. But if a composer chooses the space represented by Figure 7, I am sure there are innumerable and radically new ways to use and extend it. The future is open.

My second aesthetic step was discussed above; I list it here only for completeness.

Aesthetic Claim 2: The best music arises from an alliance of a compositional grammar with the listening grammar.

This claim does not exclude the artifice hidden in a Bach fugue or a Brahms intermezzo. Such artifice is rooted in the bedrock of a “natural” compositional grammar. At our present musical juncture, however, composers would do well to heed the claim.

This claim carries with it a historical implication. The avant-gardists from Wagner to Boulez thought of music in terms of a “progressivist” philosophy of history: a new work achieved value by its supposed role *en route* to a better (or at least more sophisticated) future. My second aesthetic claim in effect rejects this attitude in favour of the older view that music-making should be based on

"nature". For the ancients, nature may have resided in the music of the spheres, but for us it lies in the musical mind. I think the music of the future will emerge less from twentieth-century progressivist aesthetics than from newly acquired knowledge of the structure of musical perception and cognition.

References

- Babbitt, M. (1965) The structure and function of musical theory. *College Music Symposium*, 5, 10–21. (Reprinted in 1972 in *Perspectives on contemporary music theory* (ed. B. Boretz and E.T. Cone), Norton, New York.)
- Balzano, G.J. (1980) The group-theoretic description of 12-fold and microtonal pitch systems. *Computer Music Journal*, 4, (4), 66–84.
- Balzano, G.J. (1982) The pitch set as a level of description for studying musical pitch perception. In *Music, mind and brain* (ed. M. Clynes). Plenum, New York.
- Bharucha, J.J. (1984a) Anchoring effects in music: the resolution of dissonance. *Cognitive Psychology*, 16, 485–518.
- Bharucha, J.J. (1984b) Event hierarchies, tonal hierarchies, and assimilation: a reply to Deutsch and Dowling. *Journal of Experimental Psychology: General*, 113 (3), 421–5.
- Boulez, P. (1954) *Le marteau sans maître*. Universal Edition, London.
- Boulez, P. (1963) *Penser la musique aujourd'hui*. Schott, Mainz. (Tr. 1971 by S. Bradshaw and R.R. Bennett. Harvard University Press, Cambridge, Mass.)
- Boulez, P. (1964) Alea. *Perspectives of new music*, 3 (1), 42–53.
- Bruner, C.L. (1984) The perception of contemporary pitch structures. *Music Perception*, 2 (1), 25–40.
- Burns, E.M. and Ward, W.D. (1982) Intervals, scales, and tuning. In *The psychology of music* (ed. D. Deutsch). Academic Press, New York.
- Cherry, E.C. (1953) Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*, 25, 975–9.
- Chomsky, N. (1957) *Syntactic structures*. Mouton, The Hague.
- Chomsky, N. (1965) *Aspects of the theory of syntax*. MIT Press, Cambridge, Mass.
- Deliège, I.S. (1985) Les règles préférentielles de groupement dans la perception musicale. Unpublished Ph.D. thesis. Université de Liège.
- Deutsch, D. (1982) The processing of pitch combinations. In *The psychology of music* (ed. D. Deutsch). Academic Press, New York.
- Deutsch, D. and Feroe, J. (1981) The internal representation of pitch sequences in tonal music. *Psychological Review*, 88, 503–22.
- Dowling, W.J. (1972) Recognition of melodic transformations: inversion, retrograde, and retrograde-inversion. *Perception and Psychophysics*, 12, 417–21.
- Dowling, W.J. and Harwood, D.L. (1986) *Music cognition*. Academic Press, New York.
- Fodor, J.A. (1983) *The modularity of mind*. MIT Press, Cambridge, Mass.
- Francés, R. (1958) *La perception de la musique*. Vrin, Paris.
- Fux, J.J. (1725) *Gradus ad parnassum*. Joannis Petri van Gehlen, Vienna. (Tr. and ed. 1965 by A. Mann, Norton, New York.)
- Hindemith, P. (1937) *Unterweisung im tonsatz*, Vol. 1. Schott, Mainz. (Tr. 1942 by A. Mendel. Associated Music Publishers, New York.)
- Hyde, M. (1980) The roots of form in Schoenberg's sketches. *Journal of Music Theory*, 24 (1), 1–36.
- Koblyakov, L. (1977) P. Boulez, "Le Marteau sans Maître," analysis of pitch structure. *Zeitschrift für Musiktheorie*, 8, (1), 24–39.
- Krumhansl, C.L. (1979) The psychological representation of pitch in a tonal context. *Cognitive Psychology*, 11, 346–74.
- Krumhansl, C.L. (1983) Perceptual structures for tonal music. *Music Perception*, 1 (1), 28–62.
- Lerdahl, F. (1985) Théorie générative de la musique et composition musicale. In *Quoi? quand? comment?: la recherche musicale* (ed. T. Machover). Christian Bourgois, Paris.
- Lerdahl, F. (1987) Timbral hierarchies. *Contemporary Music Review*, 1, (3–4), 135–60.
- Lerdahl, F. and Jackendoff, R. (1983) *A generative theory of tonal music*. MIT Press, Cambridge, Mass.
- Longuet-Higgins, H.C. (1978) The perception of music. *Interdisciplinary Science Reviews*, 3, 148–56.
- McAdams, S. and Bregman, A. (1979) Hearing musical streams. *Computer Music Journal*, 3 (4), 26–43. (Reprinted in 1985 in C. Roads and J. Strawn, eds., *Foundations of computer music*. MIT Press, Cambridge, Mass.)
- Messiaen, O. (1944) *Technique de mon langage musical*. Alphonse Leduc, Paris.

- Meyer, L.B. (1973) *Explaining music*. University of California Press, Berkeley.
- Miller, G.A. (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, **63**, 81–96.
- Miller, G.A., Galanter, E. and Pribram, K. (1960) *Plans and the structure of behaviour*. Holt, Rinehart and Winston, New York.
- Narmour, E. (1977) *Beyond Schenkerism*. University of Chicago Press.
- Neisser, Ulric. (1967) *Cognitive psychology*. Prentice-Hall, Englewood Cliffs, NJ.
- Palmer, C. and Krumhansl, C.L. (1987) Independent temporal and pitch structures in perception of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance* **13**, 116–126.
- Partch, H. (1949) *Genesis of a music*. University of Wisconsin Press, Madison. (Republished 1974 by Da Capo, New York.)
- Plomp, R. and Levelt, W.J.M. (1965) Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, **38**, 548–60.
- Rosch, E. (1975) Cognitive reference points. *Cognitive Psychology*, **7**, 532–47.
- Rosen, C. (1972) *The classical style*. Norton, New York.
- Schenker, H. (1935) *Der Freie Satz*. Universal Edition, Vienna. (Tr. 1979 by E. Oster. Longman, New York).
- Schoenberg, A. (1911) *Harmonielehre*. Universal Edition, Vienna. (Tr. 1978 by R.E. Carter. University of California Press, Berkeley.)
- Schoenberg, A. (1941) Composition with twelve tones. In A. Schoenberg, *Style and Idea*. St Martins, New York, new expanded edn 1975.
- Schoenberg, A. (1954) *Structural functions of harmony*. Norton, New York.
- Shepard, R. (1982) Structural representations of musical pitch. In *The psychology of music* (ed. D. Deutsch). Academic Press, New York.
- Simon, H.A. (1962) The architecture of complexity. *Proceedings of the American Philosophical Society*, **106**, 467–82. (Reprinted 1972 in H.A. Simon, *The sciences of the artificial*. MIT Press, Cambridge, Mass.)
- Sloboda, J.A. (1985) *The musical mind*. Oxford University Press, Oxford.
- Terhardt, E. (1974) Pitch, consonance, and harmony. *Journal of the Acoustical Society of America*, **55**, 1061–9.
- Terhardt, E. (1978) Psychoacoustic evaluation of musical sounds. *Perception and Psychophysics*, **23**, 483–92.
- Todd, N. (1985) A model of expressive timing in music. *Music Perception*, **3** (1), 33–57.
- Weber, G. (1824) *Versuch einer Geordeneten Theorie*. Schott, Mainz.
- Werts, D. (1983) A theory of scale references. Unpublished Ph.D. thesis. Princeton University.
- Yasser, J. (1932) *A theory of evolving tonality*. American Library of Musicology, New York.