Musical memory, complexity, and Lerdahl’s cognitive constraints

John CROFT

submitted for the degree of
Master of Music

Department of Music
The University of Sheffield

Submitted in
February 1999
What is it to remember music? How important is memory, or memorability, to the enjoyment of music (or to whatever activity is deemed to be aesthetically valuable)? We often speak of comprehending music, and prima facie it would seem that musical comprehension, whatever else we might say about it, requires at the very least the ability to remember music, in order to form relations between different passages, to follow harmonic progressions, and to do all the other things that are surely essential to competent listening. More controversial perhaps, but still commonplace, is talk of musical “communication”—a process which must involve comprehension, and therefore memory, on the part of both sender and receiver.

What follows is largely concerned with the concepts of memory and comprehension in relation to the music of this century which has abandoned tonality, or metre, or both. I shall use the term ‘atonal-ametric’ to refer to archetypally atonal and ametric music; that is, music in which metre and tonality play no structural role. This term is only a convenience: there is of course much music which, though atonal or ametrical, retains vestiges of tonality or metre; and I intend the term ‘atonal-ametric’ to refer also to this music to the extent that it abandons tonality or metre.

The question, put bluntly, is: Is atonal-ametric music more difficult to remember and understand than tonal-metric music? No doubt, for many people who spend much of their time in contact with post-tonal music, the answer would be an obvious ‘no’. Others might say that it may be more difficult for those unfamiliar with it, but that that is just because it is unfamiliar. Yet others would perhaps concede that it is, in some sense, inherently harder to understand, but that this is an unavoidable aspect of music in an age when just about everything is more complex, and that it is worth the extra effort. Finally, some might argue that comprehension is not really the primary, or even a major, aim of music after all, and that the question is therefore of little importance.

On the other hand, a positive answer to this question—that is, the view that atonal-ametric music is indeed more difficult to remember and to understand than tonal-metric music—has not infrequently been used to support an aesthetic devalorisation of atonal-ametric music. It is easy, and perhaps tempting, to dismiss such views as reactionary; indeed, the view in question is often stated without argument as a kind of knee-jerk reaction on the part of those who think, as is fashionable in some quarters, that “accessibility” should be the most important concern of composers. But simply dismissing a view, reactionary or not, is an unsatisfactory response if the arguments for that view have a prima facie cogency, which I suggest is the case with at least some proponents of this view. What is more, there is always the danger that the conclusions of these writers might be used as ammunition by less subtle thinkers (and worse, by less subtle thinkers with their hands on the purse-strings of arts funding). So, even if the responses were obvious, they would be worth stating; however, the responses that I have in mind do not seem to me to be entirely obvious.

First, I should characterise the view more precisely. It involves not only the proposition that atonal-ametric music is more difficult to remember and understand, but also the proposition that it is so because there is something about atonality and
ametricality which in some sense does not “fit” with something about the human mind. Moreover, that something-about-the-human-mind is innate, and not a deficiency that could be addressed by education or exposure. My purpose is not merely to launch an attack on this view, but to take the texts of one of its more ardent proponents as an inroad to more general issues about the differences between atonal-ametric and tonal-metric music in relation to memory and cognition.

It so happens that one of the most ardent proponents is also one of the most coherent and initially plausible: Fred Lerdahl. Of course, the general view has a history, propagated by the likes of Paul Hindemith, Leonard Bernstein, and George Rochberg. Bernstein, for example, writes:

In any case, it was soon to become clear that free atonality was itself a point of no return. It seemed to fulfil the conditions for musical progress: it seemed to continue the line of romantic expressivity in a subjective way, from Wagner and Brahms through Bruckner and Mahler; the expressionism seemed logical, the atonality inevitable. But then: a dead end. Where did one go from here, having abandoned all the rules? For one thing, the lack of constraints and the resulting ungoverned freedom produced a music that was extremely difficult for the listener to follow, in either form or content. This remained true in spite of all the brilliant and profuse inner structures that abound in a piece like *Pierrot Lunaire*—canonic procedures, inverted phrases, retrogrades, and the like. Moreover, it was not easy for the composer to maintain his atonality, because of that innate tonal drive we all share universally.²

Here Bernstein expresses the view that a sense of tonality is somehow innate: try as we might to abandon or transcend it, we cannot escape the natural influence it has over the way we listen. But what exactly does Bernstein mean when he writes that we all share an ‘innate tonal drive’? It seems to me that such a claim may be interpreted in two quite different ways. One the one hand, it may amount to the claim that, as a result of the way the brain works, we tend to seek structures exhibiting, for example, goal-orientation and hierarchical organisation. On the other hand, it may be understood as the much stronger claim that the rules of tonality, or constraints which determine the set of possible rules, are specifically hard-wired in the brain, perhaps in much the same way as Chomsky considers constraints on the rules of natural language to be innately cognized. This distinction will turn out to be crucial to my discussion in the following sections.

While Bernstein focuses on tonality, George Rochberg presents a similar view with the emphasis on metre. He writes that

...the pulse-trains which transmit messages in the central nervous system suggest a direct correspondence with the logic of musical events characterized by structural continuity based on self-perpetuating forms of repetition and recall. The fact that these pulse-trains function periodically suggests that music itself may be a direct expression or reflection of the fundamental language of the human nervous system. Periodicity of

---

1 Hindemith (1945) expounds a theory based more on an idiosyncratic understanding of acoustics than on a view of the human mind: however, the rejection of atonality and the avoidance of triadic harmony as being contrary to ‘Nature’ makes sense only if the acoustical properties upon which Hindemith bases his theory are held not only to be more “natural” in themselves, but to be perceived by the mind as such. Of course, the naturalness of certain acoustical properties has continued to play a role in more recent theories which are based explicitly on properties of the mind.

metric or rhythmic movement, and the larger periodicities and rhythms of the musical phrase and of formally articulated structure, were fundamental identifying characteristics of Western art music until only very recently. Perhaps we have here a direct clue to one of the basic reasons for the essential failure of much new music to communicate itself directly to even the most sympathetic listeners. Its insistence on suppression of the pulse, its conscious avoidance of periodicity on all levels of structure and movement, and its consequent inability to perpetuate itself as a growing, identifiable, organic structural entity apparently go against the grain of the natural functions of the central nervous system.3

An important aspect of Rochberg’s view is that the absence of periodic structure make the music difficult, or perhaps impossible, to remember:

This decreasing profile of identity could be graphed in a rough sort of way, moving from a music with precise identities (Bach, Haydn, Mozart, Beethoven, Schubert, Schumann, Brahms, Wagner, Bruckner, Verdi, Strauss, Mahler, early Schoenberg) to a music with a marked decline in its profile of identity (the atonal and twelve-tone works of Schoenberg, Webern, Berg, late Scriabin, Ives) to a music entirely lacking in any aurally meaningful, identifiable characteristics (e.g. post-Webern serialist works of Boulez and Stockhausen among others; works of Cage, Feldman, Brown, based on a variety of aleatory approaches; recent works of Elliott Carter, who in an interview, expressed concern that his music cannot be remembered). In short, from a music that can be remembered, to a music which can be remembered but with varying degrees of difficulty, and finally to a music which utterly (or almost) defies memory.4

What is it about the absence of tonality or metre that, according to these writers, defies memory and understanding? While Bernstein is openly influenced by superficial analogies between linguistic syntax and musical structure, and Rochberg by psychological terminology (if not by anything recognisable as actual psychology), it is to Lerdahl that we must turn for an argument with a degree of prima facie plausibility. Lerdahl presents one of the most recent and least mysterious versions of the general idea expressed by Bernstein and Rochberg, a view we may term ‘tonal-metric nativism’.5 Lerdahl argues that in order for musical cognition to function fully, music must be organised according to certain structural principles, which he sets out in the form of a list of “constraints” upon compositional systems.6 He also argues that these required properties are absent in much recent music; and that they are required, not because of anything we have or have not acquired by learning, but because of innate and unalterable structures in the brain. Lerdahl makes no secret of the fact that the conjunction of these properties amounts, broadly speaking, to tonality and metre.

An obvious question arises here: what does it mean to say that certain types of structure are required? Surely we are not dealing with a logical necessity here, so presumably such a claim must amount to some kind of hypothetical imperative: in

---

3 Rochberg, 1984, p. 228.
4 Rochberg, 1984, p. 49.
5 I should stress that it is only the idea of the innateness of tonality and metre, and the resulting unmemorability and incomprehensibility of atonal-ametric structures, that I am claiming unites these various writers. I do not, for instance, wish to attribute to Lerdahl the excessive enthusiasm for direct analogies between linguistic syntactic structures and musical structures that characterise Bernstein’s musings on this subject.
6 Lerdahl, 1988, passim.
other words, if the claim that certain constraints upon musical structure are “binding”, as Lerdahl puts it,7 is to have any content, we must be able to say what happens when the constraints are not observed. In other words, what does abiding by the constraints buy us? While it may seem quite pointless to consider in more detail the nature of the constraints before we know the answer to this question, that is nonetheless what I propose to do: for it is only after considering the constraints that I will be in a position to say why Lerdahl’s answer is unsatisfactory. For the meantime, then, suffice it to say that, in Lerdahl’s view, without the appropriate structures we find it impossible to ‘learn and remember structure from musical surfaces’.8 The noncompliance with various constraints inhibits the inference of structure,9 and thus to incomprehension on the part of the listener.

I turn now to a more detailed consideration of the structural principles which, according to Lerdahl, ensure memorability and comprehensibility, deferring for the moment a more critical consideration of these notions.

II

We can trace the tonal-metric innateness thesis along two paths in Lerdahl’s texts, the relationship between which is not always clear: Lerdahl never quite gets around to confronting explicitly the connection between them, though they are often difficult to separate.

The first of these paths emerges from the idea of musical universals which appears in A Generative Theory of Tonal Music.10 In the Rule Index, Lerdahl and Jackendoff mark with an asterisk those rules which are specific to classical western tonality, commenting that they ‘believe the rest are universal’.11 Thus, beyond the obvious differences between the many musical idioms, past and present, western and non-western, there is a set of underlying principles which remains invariant. Moreover, Lerdahl and Jackendoff argue that it is implausible to suppose that an individual could learn a set of rules of such complexity from the various exemplars that she is exposed to. Thus, they conclude that

...much of the complexity of musical intuition is not learned, but is given by the inherent organization of the human mind, itself determined by the genetic inheritance.12

These two lines of argument for innateness—the argument from universality and the argument from learnability, are familiar from the work of Noam Chomsky, whom Lerdahl and Jackendoff note as an influence. In a more recent paper, Lerdahl aligns himself strongly with Chomsky, defining his position in opposition to empiricism.

7 Lerdahl, 1988, p. 251.
8 Lerdahl, 1988, p. 239.
10 Lerdahl and Jackendoff, 1983; henceforth GTTM.
One might argue...that all poetic methods are perceptually equal and that comprehension is just a matter of exposure. This empiricist position, which assumes that the mind is a tabula rasa that learns only by association, dates back to Locke and in its modern form has been espoused in a radical form by Skinnerian behaviorism. It was the dominant philosophical and psychological view in the 1940s and 1950s, adopted among composers explicitly by Babbitt and implicitly by the European serialists. (...) This view has since been powerfully challenged among philosophers and psychologists, beginning with Chomsky’s (1959) review of Skinner’s Verbal Behavior. Although there is ongoing debate within the cognitive and neural sciences regarding the structure and development of learning and behaviour, no one proposes going back to the assumption of an unstructured brain that is merely conditioned. ...

GTMM falls within this more recent cognitivist tradition. It is akin to linguistic and visual theories in providing a comprehensive yet detailed set of hypotheses about the structure of a mental module. 13

Both this passage and GTTM itself give the appearance of being as close as one might wish to a musical version of Chomsky’s theory. However, there is considerable ambiguity, both in GTTM and in Lerdahl’s subsequent work, which leads me to question whether this is the right way to interpret Lerdahl’s claims about innateness. Of course, in the passage just quoted Lerdahl explicitly aligns himself with Chomsky and the “cognitivist tradition”, and explicitly rejects the “tabula rasa” of the empiricists. It is all the more surprising, therefore, when Lerdahl presents, seemingly unwittingly, a series of classically empiricist arguments of the kind that have traditionally been used against Chomsky. But in order to demonstrate this I must first clarify the difference between this “recent cognitivist tradition” and empiricism, for Lerdahl’s exposition leaves a lot to be desired.

It is no accident that Chomsky is happy to refer to his view as a contemporary version of Descartes’s theory of “innate ideas”, 14 or to his innate schematisms as forms of cognition or knowledge.

13 Lerdahl, 1997, p. 422. ‘Module’ is Fodor’s term for encapsulated, dedicated, and hard-wired mental structures in which is represented information relevant to the processing of a specific kind of input. The fact that this term is used freely by many theorists should not be taken that the modularity thesis is generally accepted. Dennett, for instance, has expressed the view that Fodor’s modules are quite implausible:

Fodor’s modules are a bureaucrat’s dream: their job descriptions are carved in stone; they cannot be enlisted to play novel or multiple roles; and they are “cognitively impenetrable”—which means that their activities cannot be modulated, or even interrupted, by changes in the “global” informational states of the rest of the system. According to Fodor, all the really thoughtful activities of cognition are nonmodular. Figuring out what to do next, reasoning about hypothetical situations, restructuring one’s materials creatively, revising one’s world view—all of these activities are performed by a mysterious central facility....

By giving his central facility so much to do, and so much non-modular power with which to do it, Fodor turns his modules into very implausible agents, agents whose existence only makes sense in the company of a boss agent of ominous authority (...). Since one of Fodor’s main points in describing modules has been to contrast their finite, comprehensible, mindless mechanicity with the unlimited and inexplicable powers of the nonmodular center, theorists who would otherwise be receptive to at least most of his characterization of modules have tended to dismiss his modules as the fantasies of a crypto-Cartesian. (Dennett, 1993, pp. 260–1).

According to Dennett, one of the major problems of modularity is the problem of how (in what “format”) the output of these automata gets presented for “central processing”.

It should also be noted that the innate “knowledge” of a grammar does not imply modularity—an innately known grammar might be used only by “general intelligence” (see Fodor, 1983, p. 22).

...[W]e cognize the grammar that constitutes the current state of our language faculty and the rules of this system as well as the principles that govern their operation. And finally we cognize the innate schematism, along with its rules, principles and conditions.

In fact, I don’t think that “cognize” is very far from “know” where the latter term is moderately clear, but this seems to me a relatively minor issue, similar to the question whether the terms “force” and “mass” in physics depart from their conventional sense (as they obviously do). 15

The idea that the innate component of grammar is akin to what we normally call ‘knowledge’—that it is in some sense rule-like or propositional—is essential to make Chomsky’s innateness hypothesis non-trivial. Without it, Chomsky’s position would collapse into the trivial view that we in fact select the correct grammar. But we already know that something about the mind leads us successfully to learn language. What distinguishes Chomsky’s view from what everyone knew already is the claim that what-it-is about the mind is an innately cognized set of rules.

A good way to understand this point is to look at the arguments that have been raised against Chomsky’s view. Hilary Putnam, for example, argues that if there is a unique simplest set of rules that can be inferred from the data available to the learner, then this alone would explain the selection of the “correct” grammar, 16 and no innate schematism need be posited. 17 Thus, for Chomsky’s argument from the convergence of grammars to succeed, he must show that there is a plurality of grammars that might have been constructed, had it not been for the innate schematism. 18 A strand running through Putnam’s arguments against Chomsky’s nativism is the idea that there is nothing special about language that necessitates the postulation of a hard-wired grammar. This is why Putnam invokes a preference for simplicity as a general property of computational systems: if there is a unique simplest grammar, then it will be preferred by any generalised computational system, not just by a specialised language system: thus no such specialised system need be posited in order to explain the selection of that grammar. 15

15 Chomsky, 1980, pp. 69–70.
16 Of course, if we reject the idea of explicitly represented rules in the mind (that could in principle be read by scanning the brain), then, as Quine points out (Quine, 1974), there will be indefinitely many extensionally equivalent grammars. We may, however, understand the idea of correctness in a purely behavioural sense as follows. Imagine that a language learner has been presented with a certain number of samples of a language, and that there are at least two grammars which describe the samples so far presented, but which are not extensionally equivalent, since one counts a certain previously unheard sentence as well-formed, while the other does not. So long as there is at least one such sentence there is a behavioural criterion for distinguishing the two grammars. The set of grammars extensionally equivalent to the grammar that makes correct (which here can be understood as meaning ‘intersubjectively acceptable’) predictions for all previously unheard sentences will be correct. Quine uses the possibility of extensionally equivalent grammars to support the claim that Chomsky’s hypothesis of an internalised grammar is without empirical content. Chomsky replies that there are many sentences upon whose well-formedness or otherwise speakers agree, even though this is not decided by the data that they have been exposed to before the first presentation of that sentence. (See, for example, Chomsky, 1980, pp. 42–45 and passim.) While it is difficult to see how this response can have any bearing on the question of extensionally equivalent grammars, it does seem to show that Chomsky’s theory has some empirical content: namely, that even if there are other grammars extensionally equivalent to Chomsky’s, it is nonetheless the case that the claim that the set of grammars extensionally equivalent to Chomsky’s grammar is preferred to every other set of grammars which describes the sentences heard during language learning but which diverges on the question of the well-formedness of sentences not yet heard or uttered. Putnam’s objection still holds, however; for in order to justify the positing of an innate specialisation Chomsky must show that the set of extensionally equivalent grammars that is correct (intersubjectively accepted) does not contain a member that is simpler than any member of any extensionally non-equivalent (“incorrect”) grammar.

18 For Chomsky’s discussion of Putnam’s argument, see Chomsky, 1972 pp. 179–185.
My purpose here is not to assess Chomsky’s nativism about language, but to illustrate the distinction between the views traditionally known as Rationalism and Empiricism. This diversion is necessary because Lerdahl, while recognising these two positions as diametrically opposed, appears to misunderstand the fundamental difference between them. Remember that Lerdahl regards his view as falling within a ‘more recent cognitivist tradition’ which emphasises the importance of innate mental structures; and that he sets up an opposition between his view and the implicit behaviourist view of the avant-garde, who, supposedly, regarded the mind as a *tabula rasa*. But in fact no one ever seriously suggested that we have an ‘unstructured brain that is merely conditioned’, as Lerdahl puts it. Even the most ardent Lockean would recognise that an unstructured brain would simply not work. Nor has anyone suggested that there are no innate constraints on what the brain is capable of. The point of contention between Rationalism and Empiricism is not whether the mind is unstructured and unconstrained, but whether that structure is rule-like or quasi-propositional—and thus whether the constraints are due to general limitations, or to specialised innate knowledge. By ‘general limitations’ I mean the kind of limitations which are a feature of any computational system—such as limitations on processing speed and memory. This can be illustrated with an example from Fodor, who introduces the term “Neocartesian” to describe the kind of nativist position that he and Chomsky have adopted—a usage which I shall follow.

Suppose one believes the doctrine of George Miller’s famous paper about the ‘magical number seven’ ([19]…). Roughly, the idea is that there is a fairly constant limit on the number of unfamiliar, unrelated items that one can cope with in a task that demands immediate recall. (...) Now, one can imagine a Neocartesian treatment of this phenomenon along the following lines: there is a certain mentally represented proposition to which one gives tacit assent—viz., the proposition that, when presented with a list of *n* things to learn, one should indeed learn the first seven and thereupon forget about the rest. (...) I said that it is possible to imagine a Neocartesian story that runs along those lines, but I doubt any Neocartesian would take it seriously; and I’m sure that nobody else would. The sort of treatment that Miller’s data cry out for is not the postulation of an innately cognized rule but rather of a psychological mechanism—a piece of hardware, one might say—whose structure somehow imposes limitations upon its capacities.

... The point is ... to emphasize a distinction between two quite different accounts of what mental structures—endogenous or otherwise—might be like; one account elaborated around a notion of propositional content and the other around the notion of a psychological mechanism.20

Now it is obviously not the case that Empiricists think that the mind is unconstrained in the “psychological mechanism” sense—no one has suggested that we might learn to have an infinite memory, for example. The difference between constraints of this kind, and the constraints posited by Fodor and Chomsky, is that the latter have the character of domain-specific rules. Fodor emphasises that his modular account only makes sense if the structure of a module that gives rise to

---

19 Miller, 1956
constraints is a rule-like structure. Fodor, like Chomsky, points out that even if we avoid calling it ‘knowledge’, the fact that the role of this structure is to apply rules that we have not acquired implies that the structure must represent the kind of thing that can be known—that could, in principle, have been learned; that is, it must be propositional. The limitations of mechanisms common to all computational systems—such as memory and speed—are indeed innate limitations, but innate limitations which Locke himself would have been happy to accept.

Having determined the difference between two kinds of innate constraints—the kind that anyone would admit to, and the more controversial, domain-specific, quasi-propositional, Chomskian or Fodorian kind—we must now return to the question of what kind of constraints on musical cognition Lerdahl is arguing for. I must emphasise that what is at issue here is not whether Neocartesianism or empiricism is the right view of how the mind works. Rather, I want to ask: given that some capacities are best understood in Neocartesian terms, is the musical capacity likely to be one of them? Even assuming Neocartesianism turns out to be correct, one thing is absolutely certain: there are some capacities for which the empiricist approach is the only plausible one. There is no chance whatsoever that the Neocartesian line will be the correct one to take about the capacity for playing chess, or driving a car, or even for recognising written words—all capacities which humans acquired after they stopped evolving. In the passage quoted near the beginning of the previous section, Lerdahl comes dangerously close to arguing from the success of the Neocartesian approach in other domains to its applicability to musical cognition. However, we must consider the explanatory power of such a model separately for each domain of human competence.

What, then, are the data about musical cognition that supposedly needs to be explained by an innate schematism or module? By analogy with Chomsky’s work, we might expect two kinds of argument in particular: the first from the complexity of the musical grammar compared to the poverty of the sample from which it is inferred, and the second from musical universals; and as I have already mentioned, these are indeed the arguments found in GTTM. I shall turn to the question of universality shortly: for now I want to consider the first kind of argument, which makes an appearance when Lerdahl and Jackendoff write:

If the rules we have proposed correspond at all closely to principles unconsciously known and used by the experienced listener, one must ask how the listener manages to learn them. And of all the possible organizations one could attribute to tonal music (including all the incorrect ones posited by us music theorists), why does the listener

---

infer the one he does? The only answer that we find defensible is that one does not have to learn the entire grammar from scratch.\textsuperscript{22}

This passage strongly resembles Chomsky’s argument, suggesting as it does that some “unconsciously known” principles do not have to be learned “from scratch”. However, the passage continues:

Rather, one has no choice about much of it; many aspects of the grammar are simply the only (or easiest) ways that one’s mental abilities make available for organizing a musical signal.\textsuperscript{23}

This last remark resembles Putnam more than Chomsky. If there is a unique simplest grammar that fits the data, then, as we have seen, an innate schematism is not required to explain the selection of one possible grammar rather than another that fits the available sample. Perhaps Lerdahl and Jackendoff do not wish to commit themselves to a Chomskian form of innateness after all. But if that is the case, what is one to make of this passage from Jackendoff?

Lines of evidence from the structure of numerous languages, from historical changes in languages, from the character of child language acquisition, and from the linguistic deficits due to brain damage all converge on the view that there is a highly specified innate basis (“Universal Grammar”) from which children develop an adult language capacity during the first ten or twelve years of life. This innate basis is not just a consequence of having a large brain—of being “smarter than animals”; it seems to be a specific brain adaptation, specialized to deal with this particular form of information we call language.

The musical capacity (Lerdahl and Jackendoff [1983]; chapter 7) also appears to be an independent module (or set of modules) in the human mind.\textsuperscript{24}

We have already noted Lerdahl’s retrospective insistence that GTTM describes a mental module. Perhaps, in the light of these later texts, we should not place much weight on the empiricist moment in GTTM. Assuming, therefore, that what Lerdahl and Jackendoff have in mind is an innate Neocartesian structure for music, just how plausible is their claim?

As I pointed out above, Lerdahl and Jackendoff’s arguments parallel those put forward by Chomsky in favour of an innate schematism for language. I do not propose to assess these arguments as applied to language, but I would suggest that even if we are convinced by Chomsky’s arguments for the linguistic case, we have good reason for rejecting their musical counterparts. Chomsky argues that an innate schematism appears to be the only plausible explanation of the fact that infants learn the grammar of their native language at a very early age from an impoverished and noisy sample of it. The situation for musical competence, however, is rather different. In contrast to their competence in their native language, three-year-old children cannot understand and generate an indefinitely large number of well-formed pieces of tonal music. Tests carried out by Serafine, for example, suggest that even an ability as basic as the

\textsuperscript{22} Lerdahl and Jackendoff, 1983, p. 281.
\textsuperscript{23} \textit{ibid}.
\textsuperscript{24} Jackendoff, 1992, p. 71.
identification of straightforward tonal closure develops gradually, becoming reliable only by the age of 10, yet continuing to improve after that age. This is in marked contrast to the sophistication of language use even by the age of 3 or 4 years—younger even than the youngest of Serafine’s subjects. Another test, in which subjects were called upon to identify the underlying structure of a passage, suggests that this ability develops even later than the identification of closure. Dowling and Harwood point out that children develop the ability to use the fact that a melody is tonal in order to discern interval alterations by about age 8, and remark that ‘tonal scale schemata are among the last to form in development; among the earliest are schemata for melodic contours.’

These data suggest that the earliest musical abilities involve the perception of acoustical properties, and that the use of conventional schemata such as tonality develops somewhat later, and quite gradually. By contrast, a characteristic of innate dedicated structures is that they are activated relatively suddenly. This stands to reason: if the information is already there, there is no reason it should not be activated more or less all at once; indeed, it is the rapidity of language acquisition that the idea of innate grammar is largely meant to explain. In contrast to this, the evidence of musical development suggests that musical ability develops gradually, beginning with features that are more clearly connected with acoustic and temporal properties, while the more “grammatical” abilities develop somewhat later and gradually. While this may be counted as evidence for modules dealing with, say, temporal articulation, it appears to be evidence against specifically musical modularity. Of course, this does not prove that there is no innate musical grammar—it is not a condition of innate structures that they be put to use by a certain age—but it does suggest that the innateness thesis for musical grammars does much less explanatory work than its linguistic counterpart.

What of the other well-known Chomskian argument, the argument from linguistic universals? Chomsky urges that the existence of certain principles common to all languages strengthens the case for innateness. Are there structural similarities between the musics of different cultures that would support a parallel argument for music? Not just any universal feature will suffice for such an argument to work: in order to justify an explanation in terms of innate cognitions, the universal features need to be of reasonable complexity and somewhat “unobvious”. For example, suppose it were found that all cultures made chairs. It would clearly be absurd to attribute this universal to a specialised mental structure for chair construction, because the idea of a chair is just too obvious to need a special structure. I think we can say the same of ideas such as regular pulse, pitch centricity, and scales. Consider the example given in the context of Lerdahl and Jackendoff’s version of the universality argument:

The range of variation among rules in different idioms also constitutes grounds for hypotheses about innateness. For example (to consider an extreme case), though idioms differ in metrical and intervocalic possibilities, we feel safe in conjecturing that there is

---

26 Serafine, 1988, pp. 213ff.
no idiom that makes use of metrical regularities 31 beats apart, or for which the most stable melodic interval is the thirteenth.\textsuperscript{29}

The trouble with this is that such similarities are just too obvious to call for explanation in terms of an innate grammar or module. The preference for somewhat smaller, more readily singable, intervals and simpler divisions of musical time are just the most obvious musical structures, just as chairs are the obvious things to sit on. This is really another version of Putnam’s point: if a structural principle is simpler than the alternatives, it does not need to be explained by innate cognition.

\textit{\textit{\textit{\textit{iv}}}}

Finally, it has to be said that the development of an innate module for cognizing music is extremely unlikely from an evolutionary point of view. Jackendoff recognizes this fact, but, surprisingly, does not seem to consider it to be a problem.

There appears to be no evolutionary justification for such a capacity: unlike language, it is hard to think of convincing reasons why such a capacity should be adaptive. However, the principles and representations organizing the cognition of music are in part peculiar to music, and all of Fodor’s criteria for input modules apply to musical perception. Thus it appears that one must acknowledge a separate specialization in this capacity as well.\textsuperscript{30}

This issue is also mentioned by Lerdahl and Jackendoff in GTTM,\textsuperscript{31} where, instead of insisting on a music module, they suggest that this objection might be overcome by regarding their innate musical grammar as ‘a confluence of a number of factors independently needed for essential aspects of cognitive ability,’\textsuperscript{32} (therefore, not as a ‘separate specialisation’ as Jackendoff suggests above). This strikes me as a more plausible solution.\textsuperscript{33} The various capacities that work together in this way may of

\textsuperscript{29} Lerdahl and Jackendoff, 1983, p. 282.
\textsuperscript{30} Jackendoff, 1992, p. 71. It must be remembered that the mere fact of specialisation does not imply an innate grammar. The fact that ‘the principles and representations organizing the cognition of music are in part peculiar to music’ is not in itself an argument for innate structures. Many human activities involve principles, and presumably “representations”, peculiar to those activities. Playing chess, for example, or Morris dancing. For the existence of such specialisations to support a hypothesis of innateness (in any interesting sense), they would have to be specialisations of a kind or of a complexity that could not be acquired by learning—this returns us to the point made at the end of the previous section.
\textsuperscript{31} Lerdahl and Jackendoff, 1983, pp. 282–283.
\textsuperscript{32} Lerdahl and Jackendoff, 1983, p. 283.
\textsuperscript{33} Dowling and Harwood (1986, p. 236) appear to believe a solution to this problem is to be found by appealing to the notion of group selection. This explanation is unsatisfactory, since, in general, group selection does not work. To be passed on, a gene must be “good for” the individual (or its offspring). (More accurately, a gene needs only to be “good for” itself—the point is that a gene is more likely to be good for itself if it is good for the individual; whereas its being good for the group is irrelevant to the gene’s survival, unless it is also thereby good for the individual. (For a discussion of this point, see Ruse, 1985.) All of this is to be distinguished from the non-biological (“memetic”) evolution of ideas and practices within a culture: the meme for musical practice may very well be selected, but this is entirely different to a gene for musical ability, and only the latter can explain an innate grammar or specialisation. (See Dawkins, 1976.)
course themselves be modular, but we would no longer have a music module.\textsuperscript{34} However, if Lerdahl follows this path, then his view of the musical capacity becomes in an important sense an empiricist one: there are no dedicated structures for music; it just so happens that when listening to music we make use of a combination of other abilities, modular or otherwise. They may of course have their own limitations, but since there is nothing that can be called a *music* module, it is now an open, empirical question whether other abilities may also be called into service when cognizing music. If musical cognition borrows capacities evolved for other purposes, then all potentially useful capacities must be considered before classifying a given musical structure as incompatible with musical cognition: in other words, we must take an empiricist approach to the question. It cannot be had both ways: one cannot insist on an “innate specialisation” which is nonetheless made up of a combination of mental capacities which are borrowed to perform the task at hand, since in principle one could borrow any capacity, modular or otherwise, to do the job. The only way to save the “innate specialisation” aspect of the “confluence” view would be to define music as being necessarily based on speech, for example. Obviously this is unsatisfactory, for it would amount to an arbitrary exclusion of a large quantity of contemporary music whose structure does not fit the phrasal paradigm—but the cognitive opacity of such music must be demonstrated, not assumed.

But assume, for the sake of argument, that one did somehow justify the labelling of some concatenation of capacities as an ‘innate music specialisation’, and that the rules of GTTM were somehow found to be represented in this concatenation. Even this implausible state of affairs implies nothing at all about atonal or ametrical music. It says nothing about our ability, or inability, to appreciate music which does not engage this grammar. Even if the existence of a innate grammar for tonal-metric music were demonstrated, this would leave open the possibility that atonal-ametric music might engage other capacities in an equally interesting way. In other words, we are free to regard this specialisation, if it exists, as a tonality specialisation, rather than as a music specialisation, having little bearing on atonal or ametrical music (or, perhaps, coming into play only at certain points in such music).

This objection may seem like philosophical sleight of hand: surely, it may be replied, the purpose for which a set of innate capacities evolved does not prevent that set of capacities from constituting an innate schematism for music. Indeed, I would reply, that would be the case if “cognizing music” were a unitary activity with definite criteria for success, for then it would be quite conceivable that a particular set of innate capacities must be used in order successfully to cognize a piece of music. But it is the existence of criteria for success—not the evolutionary history of a capacity—which would allow us legitimately to label an innate capacity as a specialisation for music. The reason Chomsky’s schematism may be regarded as an innate specialisation for *language* is not simply that it evolved specifically for language; rather, it is that there exist definite criteria for successful linguistic communication. In calling his schematism a *language* schematism, Chomsky must be understood as claiming that no language which did not conform to the constraints (and which therefore would not provide valid

---

\textsuperscript{34} Though Lerdahl and Jackendoff suggest that there might be a ‘residue of musical cognition that is a product of nonadaptive evolutionary accident.’ (Lerdahl and Jackendoff, 1983, p. 283)
input for the schematism) could be learned or used successfully; in other words, we could not use such a “language” to do what we normally do with language. This amounts to a claim about fairly obvious behavioural patterns: to say that a given non-conforming “language” is unlearnable and unusable, one is claiming that it is impossible for a group of speakers effectively to make requests, give instructions, ask and answer questions, and so on, in the language. In the case of music, however, there do not seem to be such behavioural verification conditions for successful cognition. It will not do to say, for example, that a listener must be able to recognise when a theme returns, or to know where the cadences are, or anything of that kind, since to do so would be to exclude a priori all atonal or athematic music. One cannot define the innate music specialisation as that set of capacities which allow us to follow tonal music, and then argue that there is something amiss with atonal music because it fails to fit that specialisation.

This circularity would of course be present even if this specialisation were supposed to be a single module evolved for musical cognition: even then, the existence of this module would no more give rise to an imperative to produce tonal-metric music than the existence of a module for depth-perception would make perspective an obligatory feature of all paintings, or than the existence of a module for face recognition would compel painters to produce only portraits. The suggestion that the posited specialisation is likely to be a combination of capacities which did not evolve for music simply adds another level of arbitrariness: since there is no longer even the theoretical possibility of a neurophysiological or biofunctional justification for the identification of a “specialisation”, it would be their use for tonal-metric cognition alone which would justify their being considered as constituting a single capacity.

It would appear from the foregoing considerations that the musical innateness thesis does less explanatory work than Chomsky’s original linguistic version. If one needs any more convincing of this, one need only turn to “Cognitive Constraints on Compositional Systems”, in which Lerdahl explains all of his musical universals as arising from general considerations of simplicity and memorability. Given Lerdahl’s explicit endorsement of an anti-empiricist interpretation of his position, it comes as a surprise to find Lerdahl providing as close an analogue as one might wish to Putnam’s tactic against Chomsky’s nativism about language. The arguments he presents are almost entirely of the Lockean, empiricist, magic-number-7 kind; not the Neocartesian, Chomskian, or modular kind. They are based on general cognitive limitations, not on innately known rules.

Lerdahl presents his views about what properties must be true of the musical surface to facilitate musical cognition in the form of seventeen “constraints”. The logical and cognitive status of these will be the topic of much of what follows; however, a brief survey of the kind of restrictions imposed will help to drive home my point that Lerdahl in fact provides a Putnamesque argument against his own professed stance on innateness. The most significant of these involve limitations of memory. For example, Constraints 1 to 3 all stem from the general importance of hierarchical structure in perception, ultimately as a means of reducing memory load. Similarly, Constraint 4 emphasises the importance of repetition, Constraint 5 of regularity of phenomenal accent, and it is these considerations that lead to the principles of time-
span segmentation and of prolongational trees that are familiar from GTTM. One might say the same of the constraints on pitch structure, such as the requirement of a “fixed collection of elements” (Constraint 9), and the requirement that the intervals are ‘large enough for adjacent elements to be easily discernible, but not so large as to take up excessive space along the continuum’ (Constraint 10). Constraint 11 requires that the pitch collection recur at the octave. As Lerdahl writes, ‘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.’ I think it is clear from all this that whatever Lerdahl says he is doing, what he is in fact doing here is not describing a module, nor an innate grammar of the Chomskian variety, but is rather pointing out general limitations—the kind of innate constraints that Locke would have approved of.

Given this, it is perhaps difficult to see why Lerdahl introduces the concept of modularity at all—though a possible reason might be his apparent identification of modular processing with “intuitive” understanding, and “all-purpose reasoning” with “analytic comprehension”. This identification is wrong: that something “comes naturally” certainly does not imply that it has its own module—unless we are willing to posit dedicated modules for driving a car, playing the piano, juggling, and recognising individual words of written English. Conversely, the fact that an ability is based on “general intelligence” does not imply that it involves conscious or deliberate analysis, or “figuring out”, as Lerdahl describes it.

So, not only is the idea of an innate music grammar or module implausible in itself, it turns out that Lerdahl himself provides an exemplary argument against it. Whatever Lerdahl claims to be doing, he is not really arguing for an innate grammar or a music module at all, but rather for constraints arising from quite general limitations of the mind. This will become even more apparent below, when I consider some of the constraints in greater detail. For this reason, I shall put aside the idea of innate grammars and music modules, and consider Lerdahl’s apparent project in “Cognitive Constraints…” of deriving certain restrictions on musical structure from the capacities and limitations of the mind in general.

V

“Cognitive Constraints…” begins with a discussion of Boulez’s *Le Marteau sans Maître*. In this piece (and in serial music generally), we are told, it is impossible to hear the serial structure; and according to Lerdahl, the fact that we do not pick up the permutations of the series when we listen to serial music should be cause for concern. Exactly why this should be cause for concern is not always clear: one continually gets

35 This example is from Churchland, (1988) (p. 177), and is discussed by DeBellis (1995) (pp. 105ff). Fodor replies that the recognition of written words (and of automobiles) is a matter of making “educated inferences” from properties like shape, colour, and sequence (Fodor, 1988, p. 259). As DeBellis notes, this is strange given that Fodor claims that the application of “basic concepts” like ‘dog’ is modular. At any rate, we see from this that even Fodor does not think that apparent automaticity of a capacity does not in itself imply modularity.

36 Lerdahl, 1988, p. 255.
the impression that Lerdahl wants to say that this music is aesthetically inferior, but he stops short of saying that explicitly. In fact, in his efforts to downplay the normative aspect of his position, he goes so far as to write:

"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."

Despite this disclaimer, I think we must take Lerdahl’s professed opinion that serial pieces are on average as “good” as tonal music with a rather large grain of salt. Serial music cannot help but fail most of Lerdahl’s Constraints, whereas, not surprisingly, most tonal music has no trouble conforming. If the Constraints do in fact tell us how to write music which is ‘aesthetically available,’ and if ‘the best music utilizes the full potential of our cognitive resources,’ then it is difficult to avoid a normative interpretation of Lerdahl’s view. Indeed, apart from the avoidance of seeming reactionary and conservative, it is difficult to see why Lerdahl downplays the normative implications—after all, if a serial piece is as likely to be “good” as a tonal piece, then the Constraints would seem quite pointless.

Whatever we make of the normative aspect, I would argue that the inaudibility of note rows is a spurious problem, for two reasons:

1. Lerdahl misrepresents the situation by portraying the “sequence of events” as the only input to musical understanding. As Nicholas Cook writes,

   "Audibility, in short, is not everything in music. Dahlhaus writes that ‘an undogmatic theory of art must recognize that the criterion of audibility, of complete realization by perception, is not a natural law of aesthetics but a postulate of historically limited scope. By rigorously restricting the concept of music or of “music proper” to the perceptible, one curtails historical reality for the sake of a dogma not older than the eighteenth century.’ One cannot reasonably demand that music must, by definition, yield all its meaning in perception. It would obviously be narrow-minded to deny the aesthetic validity of Machaut’s palindromic chanson Ma fin est ma commencement, or to refuse to recognize it as music, simply because of the impossibility of grasping its structure in purely perceptual terms; it is equally narrow-minded to reject a piece of serial music (as people actually do, or at least used to do) on the grounds of its consciously adumbrated organization, without giving it a hearing first."

In criticising serial music for having an inaudible structure (in that the permutations of the row cannot be discerned through listening alone), Lerdahl subscribes to the view that Cook warns against in this passage—indeed, he seems to make a principle of not ‘giving it a hearing first’:

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...

---

37 Lerdahl, 1988, p. 251.
39 Lerdahl, 1988, p. 255.
40 Lerdahl, 1988, p. 234.
41 Dahlhaus, 1983, p. 54.
42 Cook, 1990, pp. 8–9.
issue here. The issue is why competent listeners do not hear tone rows when they hear serial pieces.\textsuperscript{43}

Despite Lerdahl’s professed interest in music as it is perceived, his complaints about serialism are based for the most part not on the influence of serial technique upon the heard result, but upon the mere fact that the series cannot be heard.

“Cognitive opacity” of this kind is not restricted to serial music: as Cook writes, ‘there is no evidence to suggest that Machaut or Isaac or Bach was worried if his listeners failed to perceive the contrapuntal devices in his music, any more than there is evidence that Haydn or Mozart or Beethoven worried that his listeners might fail to perceive a modulation or a recapitulation.’\textsuperscript{44} Ironically, Lerdahl allows a concern with inaudible structures to prevent an unbiased investigation of the music as it is heard—his comment that, in the case of \textit{Le Marteau}, ‘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,’\textsuperscript{45} functions more as an assumption than as a conclusion drawn from a consideration of what the music actually sounds like.

Not only can inaudible structures be meaningful and relevant in themselves, but collateral knowledge of the music, including the knowledge that it was composed using a certain technique, tends to affect the way we hear. It is surely not unreasonable to say that if, having analysed the score, I become aware of hidden structures of which I was previously unaware, then I am able, in a straightforward sense, to \textit{hear} these structures as I listen to the piece, and that this might add another level of interest to the music, albeit one that may be of less importance than those discernible without such collateral knowledge. My point is not merely that we are able to have the (admittedly limited) pleasure of knowing \textit{that} a piece we are listening to is serial; but that if we also know something about \textit{how} it is serial, then we may even come to hear a certain passage \textit{as}, say, a simultaneous exposition of the row and its retrograde inversion. If one is prepared to admit that one is more likely to hear a recapitulation if one knows a little about sonata form, then one should have no trouble admitting the pertinence of collateral knowledge in serial music as well.

In fact, I think Cook suggests too clear-cut a distinction when he writes:

I am not denying that a technical analysis may uncover aspects of the musical construction that are interesting or elegant or even, in their own way, beautiful. But we are dealing here with a musicological rather than a musical beauty.\textsuperscript{46}

That the middle movement of Berg’s \textit{Chamber Concerto} is palindromic is probably not perceptible through listening alone. But, since I know that the second half of the movement is an exact mirror-image of the first half, I do in fact \textit{hear} various melodic

\textsuperscript{43} Lerdahl, 1988, p. 251. Note also the following footnote:
Plainly [Schoenberg] 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. (Lerdahl, 1988, p. 251n.)

\textsuperscript{44} Cook, 1990, pp. 216–7.

\textsuperscript{45} Lerdahl, 1988, p. 233.

\textsuperscript{46} Cook, 1990, p. 166.
fragments in the second half as retrogrades of the first. I hear descents as retrogrades of the ascents that I remember from the first half, ritardandi as the reverse of accelerandi, and the middle as an immense upheaval whereby the momentum of the first half is reversed. These are audible features, once I know about them—but is it a case of “musical” or “musicological” appreciation? Perhaps the initial piece of collateral information (‘it is palindromic’) is musicological, but the additional appreciation of the piece as I listen is surely musical. At any rate, my point is not that the processes of discovery are the same, nor even that the type of knowledge is the same (after all, the fact that one kind of knowledge assists in the acquisition of another kind does not make them the same kind of knowledge); but rather that once you have that knowledge, there are not two types of appreciation going on in parallel.47

If paratextual information can indeed be musically relevant, then Lerdahl’s claim that ‘the listener hears the acoustic signal, not its compositional specification’48 is misleading: its plausibility depends on its propagation of a confusion between the (true) claim that the physical input to our auditory system is the signal alone, and the (false) claim that any description of what we hear is independent of what we know or expect—Lerdahl relies, ironically, on his own version of a cognitive tabula rasa.

(2) Even were we to take the musical signal as the only input to cognition, it would be a mistake to dismiss the cognitive relevance of serial technique. Lerdahl himself acknowledges that Boulez

shaped his materials more or less intuitively, using both his ‘ear’ and various unacknowledged constraints. In so doing, he listened much as another listener might.
The [serial] organization...was just a means, and not the only one, towards an artistic end.49

Lerdahl seems to acknowledge here that there is more to “cognizing” Le Marteau than deciphering the permutations of the series. In fact, I doubt that many listeners to serial music, nor many composers who have used serial techniques, have believed such an ability to be at all relevant to the appreciation of a piece.50 So why does Lerdahl suppose it to be relevant?

The problem with works such as Le Marteau, Lerdahl tells us, is that the “listening grammar” informs only the intuitive constraints, and not the “compositional grammar” (in this case, serial technique). A “compositional grammar” can only be legitimated, in Lerdahl’s view, by being ‘based on the listening grammar’.51 Lerdahl seems to rule out, without an argument, the idea of the compositional system as merely a compositional

---

47 Cook acknowledges that “musical” and “musicological” listening can interact fruitfully, while insisting that they are fundamentally different. I don’t think that the rejection of a principled distinction here affects the polemical point he is making in this passage against textbook “appreciation”. Interestingly, Cook is happy to acknowledge that in the case of semantics, the understanding we have of the work will affect what we actually perceive:

Someone who perceives Figaro as a domestic comedy will experience it in a distinctly different manner from someone who perceives it as a political document; and this does not simply mean that they will hear and see the same things but interpret them differently. It means that they will actually perceive different things... (Cook, 1990, p. 168.)

48 Lerdahl, 1988, p. 234.
50 An exception might be some of Webern’s music, in which the permutations of the row are made audible by means of reinforcement by timbral, rhythmic, and registral relations.
51 Lerdahl, 1988, p. 236.
tool, useful as a counterbalance to intuition, perhaps to generate ideas that the composer would not normally have thought of: this, for example, is Boulez’s view of such techniques.

More significantly, I think this points to a confusion at work in “Cognitive Constraints...”. Suppose we accept, for the sake of argument, that Lerdahl is right about the “listening grammar”—which, presumably, describes the hard-wired unchangeable mental structure that constrains the musical structures that we are able to understand. Even this would not justify an imperative that imperceptible compositional devices not be used; it would merely justify a requirement that such techniques not be used if they necessarily generate music which cannot be cognized. Yet we have seen that Lerdahl acknowledges that serial music may contain other structures that are quite cognizable—only to rule out ‘what listeners actually infer’ as beside the point that he is making. I should have thought that it was precisely the point, and that one ought instead to regard the point about the inaudibility of serial structures as a red herring. Of course, it may be that serial technique is more likely to generate music which fails to conform to Lerdahl’s constraints—in fact, that seems quite likely given that most serial music is atonal: nonetheless, that claim is quite separate from the point that we cannot follow note-rows when listening to serial music.

Perhaps the problem is this notion of being “based on”. What does it mean for a compositional system to be “based on” the listening grammar? Lerdahl appears to interpret this narrowly as requiring the audibility of note-rows. However, this is surely to set the resolution arbitrarily fine—why not just say that “hearing serialism” just means hearing the absence of tonal centre resulting from using every note with equal frequency and the avoidance of tonal intervals, or hearing the influence of the intervallic content of the series, or perhaps even getting an occasional impression of retrogrades? All of these are possible. To put this in terms of Nattiez’s model of musical communication, I would suggest that Nattiez’s remark that the ‘esthesis process and the poietic process do not necessarily correspond’52 is true but misleading, since the process of composition and the act of listening are just not the kind of things that could correspond. At best, the act of listening can correspond not to a poietic process, but rather to a selection of what is relevant about the result of that process. And if we admit this, then it cannot be said simply that serial structure is inaudible—for its results are audible in the senses just mentioned.

It might be replied that serial structure requires that the poietic process not be underdetermined by the heard result. For example, if I hear the intervallic influence, or

---

52 Nattiez, 1990, p. 17. Certainly it has to be said that nothing is gained from Nattiez’s reversal of the second arrow in the classical formula (giving “Producer” → Trace ← Receiver)—a move that only adds to the confusion. In the “classical” model (“Producer” → Message → Receiver), the arrows indicate information flow, made possible by the existence of a unique encoding process that corresponds to a unique decoding process. Remove the latter, and the first arrow no longer makes sense, except to signify the material fabrication of the trace. But this is an utterly different activity to whatever might now be represented by the second, reversed arrow, which does not affect the trace in anything like the sense of the classical model: there is nothing flowing from the Receiver to the Trace that could be represented by the arrow. What the Receiver actively constructs is not the trace, but an intentional object. (The composer may do this as well—this is not poiesis, however, but the composer’s aesthesis.)

Of course, Nattiez might reply that the arrows are meant only to suggest a generalised notion of an activity directed towards the trace—but this would owe any prima facie plausibility to its vagueness. Being “directed towards” in the sense of “making” is just so different to being “directed towards” in the sense of “interpreting” that the model remains useless—for that matter, one might as well turn the second arrow in the “classical” model around, because the Receiver is directed towards the message in the sense of “decoding”.

19
even the occasional retrograde, this does not determine that the piece is serial in the way that my hearing all the forms of the row would. This would be to require the cognized result to be sufficient for a certain poietic process (in other words, the poietic process would be a necessary condition of the cognized result). The obvious response to this is that not only is the result of a poietic process subject to interpretation and selection of salient features, but so is the process itself. Serial technique might instead be described as a method whereby the interval content of a passage of music is restricted, and approximately equal importance is placed on each pitch. Described in this way, the results are certainly audible—one need only compare a passage based on a row made predominantly of thirds to one based on a row constructed from fourths and semitones. These two features almost determine serialism; and for the rest—well, the intervals have to be in some order, which might as well be the same each time, but this is not an essential part of the musical style: Karel Goeyvaerts, who is often credited with having invented integral serialism, did not order his pitches and durations, but maintained a uniform distribution of them. Presumably Lerdahl would want his objection to apply to this music as well, which, after all, is in a very similar style to Boulez’s Structures Ia. This reinforces my point that the audibility of note-rows is a red herring.

This betrays a fundamental problem: Lerdahl seems to assume that, because the composer uses a particular technique, it must be some kind of “grammar”. But to characterise musical perception as involving a single box labelled “listening grammar”, which takes the “sequence of events” as input, and produces the “heard structure” as output, is either arbitrarily to confine musical experience to something resembling the parsing of strings of a language, or to use the notion of “grammar” in a hopelessly loose way. I am not sure which of these is the best way to interpret Lerdahl, but these appear to be the only options; for it is obvious that most of human experience is not usefully understood as grammatical. Take the example of natural language, the original locus of the notion of grammar: it is clear that not all facts about the production of speech sounds are parts of the grammar—it would be obviously wrong-headed to insist that a person be able spontaneously to perceive the precise way in which vowels are produced, despite the fact that vowel sounds need to be produced in some way, just as the notes need to be in some order. This mistake is analogous to interpreting the distribution and instrumentation of the notes in tonal chords as part of the grammar of tonal music. The notes have to be played by something, in some register, and indeed some arrangements of the notes in a chord will be better than others; but it would be a mistake to suggest that the octaves, doublings, and instrumentation of harmonies in tonal music ought to be ‘transparent to perception’. There is always more information in the signal than is relevant to understanding the message.

Thus, if we interpret serial technique at the right level—as a technique designed to minimise the sense of tonal centre, and to influence the harmonic and melodic

---

53 I would not want to deny, however, that in some cases the serial structure may be audible in a more specific way. In much of Webern’s music, for instance, where the pitch series are reinforced by other parameters, it is possible to hear retrograde forms. Krumhansl et al. (Krumhansl, 1987) present evidence that serial transformations in music are audible in some contexts.

54 It is strange that Lerdahl’s theory seems to necessitate such a thing as a ‘sequence of events’, since he argues in “Composing and Listening” (Lerdahl, 1997, pp. 425ff) that Nattiez’s trace level is untenable. But if that is the case, then what is this ‘sequence of events’ which exists prior to any cognition?
intervallic “colour” of the music, then we find that it is “based on” listening after all. Not only serialism, but also techniques based on geometry, probabilities, and fractals, such as those used by Xenakis, have a perceptible outcome, and are used deliberately to produce music with certain perceptible characteristics which would be more difficult to achieve by other methods. Thus, if Lerdahl’s claim is that such methods are ‘unresponsive to musical listening’,\textsuperscript{55} then it is false.

We must therefore avoid confusing the transparency of compositional techniques and the comprehensibility of the resulting musical structures. Lerdahl obscures this distinction in an interesting way. Consider, for example, the following passage.

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. ... What other foundation was there to turn to but the nature of musical understanding itself? In other words, I decided that a compositional grammar must be based on the listening grammar. ...

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).\textsuperscript{56}

Notice how Lerdahl first draws attention to the dangers of a compositional technique which is ‘unresponsive to musical listening’. The solution to this is to pay attention to ‘the nature of musical understanding itself’, which is, it seems, the same as the “listening grammar”. But help is at hand, for a description of this has already been undertaken, in the form of \textit{A Generative Theory of Tonal Music}.

The turn of phrase ‘the listening grammar’ operates as a clever pivot here, between, on the one hand, the apparently neutral idea of “musical listening”, and on the other, a set of rules describing the structure of tonal-metric music—GT TM, or, in an alternative formulation, the constraints Lerdahl is about to set out in “Cognitive Constraints...”. On a less careful reading one might be left with the impression that these are supposed to be the same thing. In fact, they are very far from being the same thing. In the previous section I pointed out that techniques such as serialism can indeed be aimed at entirely perceptible musical results, such as the absence of tonal centre and control of harmonic or melodic intervallic “colour”. Serialism is, so far as it goes, responsive to musical listening after all. But it certainly does not often produce results which have much to do with the “listening grammar” as it is described in GT TM. So, in the end, Lerdahl gets the result he wants—serial (and, as we shall see, a great deal of other) music fails to fit the listening grammar; but only at the expense of the coextensiveness of “musical listening” and “the listening grammar”.

This is the beginning of a big problem. There are perfectly perceptible musical features which have little to do with Lerdahl and Jackendoff’s description of the innately specified music cognition faculty. This would be all well and good, if Lerdahl were not about to use this fact as a cudgel against atonal-ametric music. Of course, if distinctive interval content was the limit of perceptible structure in atonal-ametric music, then this would perhaps be a minor omission from Lerdahl and Jackendoff’s

\textsuperscript{55} Lerdahl, 1988, p. 236.

\textsuperscript{56} Lerdahl, 1988, p. 236.
theory, one which Lerdahl would no doubt dismiss as insufficient to sustain a musical structure—perhaps generating attractive sounds, but little more. For despite passages such as the one just quoted, I doubt that “musical listening” and “the listening grammar” were ever meant to be the same thing. I suspect that the reason Lerdahl insists upon a narrow interpretation of what would count as hearing serial structure is that only this narrow interpretation, based as it is on the identification of row forms and note order, could turn serialism into a multiple-level hierarchical structure to rival tonality. The point is not that serialism is not informed by listening—we have seen that it is, so far as it goes (and for the avant-garde it was never really meant to go beyond the generation of consistent atonal material). Rather, the point is that serialism is not a system of audible hierarchical levels.

If this is Lerdahl’s point, then, having made it explicit, we might simply reply that we knew that already. But its very implicitness in Lerdahl’s text seems to save him the trouble of considering more likely candidates for alternatives to tonal structure. If one did not notice the hollowness of this victory over serialism, one would be left with the impression that Lerdahl has actually demonstrated something important about serialism—something which, as his paper progresses, somehow makes all other atonal-metric music guilty by association. But despite this curious opening, the constraints need to be considered in their own right, each one standing or falling according to its plausibility. Given this, it might seem even stranger that Lerdahl would begin with the pseudoproblem of serial structure—however, this allows the constraints to be presented as if they were a solution to a determinate problem: ‘the gap between compositional system and cognized result’ which has, we are told, resulted in contemporary music having ‘lost its way’.

vi

I have drawn attention to one quite audible structure—the intervallic coloration of material resulting from serial technique—which does not appear in Lerdahl’s “listening grammar”. As I have pointed out, this is not surprising given that the “listening grammar” is a description of a kind of music, not of everything that could be considered an audible structure. Yet, as I have also insisted, this exclusion is arbitrary, since the only plausible justification for such an exclusion would depend on a criterion of successful musical cognition, which itself cannot be specified without making assumptions about what is worthy of musical attention—assumptions which themselves depend upon the a priori valorisation of a particular kind of music. Of course, if harmonic and melodic intervallic colour were the extent of perceptible structure available to atonal music, then that would certainly be a victory for proponents of tonality. Indeed, the impression one gets from Lerdahl is that any modes of cognition excluded by the Listening Grammar are no threat to his position, since they

57 Lerdahl, 1988, p. 234.
58 Lerdahl, 1988, p. 236.
allow only the most impoverished cognition, a minimal engagement of our mental capacities, the vaguest mental representation of the musical surface, and a musical experience that amounts to little more than a continuous stream of possibly pleasant sounds. I want to suggest, on the contrary, that this view is based on a misleadingly narrow picture of what the mind is good at, and that these other, excluded modes of listening are at least as valid candidates for musical cognitive norms as the principles that Lerdahl puts forward. Having created the impression that there is no alternative to his “listening grammar” by appealing to the notion of an innate grammar or module and to the narrative of the gap between composition and cognition, it is much easier for him to make tacit assumptions about the kind of structures that might be relevant. I shall reveal these hidden premises and argue that there are in fact many significant ways of listening that have little or nothing to do with Lerdahl’s “listening grammar”.

In assessing Lerdahl’s constraints one is confronted with question of what kind of statements they are supposed to be. As I suggested above, the normative aspect of Lerdahl’s work is impossible to avoid; but we cannot interpret the constraints simply as the statement of an aesthetic point of view, for this would fail to account for their alleged derivation from empirical claims about the capabilities of the mind. On the other hand, many of the constraints are not themselves empirically falsifiable statements.

We can distinguish two kinds of constraint. First, there are those which include the word “must” or “should”; for example:

**Constraint 1**: The musical surface must be capable of being parsed into a sequence of discrete events.⁵⁹

Second, there are those which describe the necessary conditions for some feature to be present—usually a feature whose importance has been affirmed by a constraint of the first kind. An example is:

**Constraint 3**: The establishment of local grouping boundaries requires the presence of salient distinctive transitions.⁶⁰

We need to deal with these two types of constraint in different ways. Constraints of the second kind are empirical claims which are open to falsification; taken individually they have no aesthetic dimension. Constraint 3, for example, neither valorises nor devalorises ‘the establishment of local grouping boundaries’; it simply tells us what to do if we want them.

Any aesthetic import that constraints of this kind have derives from constraints of the first kind, which have a very different character. They appear to state categorically what “must” be the case. There are in general only three ways to interpret a sentence containing an unqualified “must” as making a claim that might be true or false. First, we can ignore it, taking it as equivalent to “in fact” (as in the folk physics saying, ‘What goes up must come down’). Second, we can take it as a modal operator of necessity (as in ‘One plus one must equal two’—that is, 1+1=2 in all possible worlds). Obviously neither of these is a plausible interpretation in the present case. Finally, we

---

⁵⁹ Lerdahl, 1988, p. 239.
⁶⁰ Lerdahl, 1988, p. 240.
can interpret it has having a hidden hypothesis: assuming you want \( q \) to be the case, you must ensure that \( p \) is the case. This is clearly the interpretation that Lerdahl intends. However, the omitted hypothetical part of these ‘must’-constraints is really quite vague: what is the \( q \) for all these statements that \( p \) must be the case? The word ‘cognition’ appears frequently in Lerdahl’s work; however, from the arguments I have already presented—that there is no encapsulated module or dedicated innate grammar, and that there is no predetermined criterion for success of musical cognition—it should be apparent that to characterise the constraints simply as the conditions for the “cognizability” or “comprehensibility” of music gets us nowhere. As for ‘cognitive transparency’\(^{61}\) and ‘transparency to perception’,\(^{62}\) it is again difficult to imagine an interpretation of these terms that relies neither on the notion of communication (which again depends on the definition of criteria for success) nor on the red herring of the audibility of compositional processes. If the constraints are to have content, we need something specific that they enable the listener to do.

It may appear that this problem is solved by Lerdahl’s insistence on the importance of memory in musical cognition. Aside from the more amorphous notions just mentioned, much of Lerdahl’s discussion focuses on the aim of making it possible to remember the music, to ‘infer structure’\(^{63}\), or to ‘assign a precise mental representation to what is perceived.’\(^{64}\) This may seem uncontroversial—surely the ability to remember music is desirable, and thus can be accepted as a criterion of aesthetic value. However, in my discussion of the constraints, I shall urge that Lerdahl uses the idea of memory in an unanalysed and simplistic way that neglects the actual function of memory in cognition in general. In treating memorability as an end in itself, he is able surreptitiously to import tonal norms as tacit \textit{a priori} assumptions, thus illegitimately excluding the possibility that quite different principles might apply to the perception of atonal and ametric music.

\[ \text{vii} \]

Constraints 1 to 4 establish the fundamental cognitive importance of hierarchical structure. It is from these first four constraints that Lerdahl, in the subsequent constraints, derives the properties of metricality and tonality as obligatory elements in music. Constraint 5, for instance, states that the ‘establishment of metrical structure requires a degree of regularity in the placement of phenomenal accents.’\(^{65}\) Constraints 7 and 8 emphasise the importance of stability conditions—hierarchical organisation requires that some events be heard as subordinate to others. This prepares the way for the introduction of tonality as an obligatory feature.

\[^{61}\] Lerdahl, 1988, pp. 237, 255.
\[^{63}\] Lerdahl, 1988, p. 239.
\[^{64}\] Lerdahl, 1988, p. 232.
Two questions thus arise. First, are the general claims about hierarchy in Constraints 1 to 4 true? Second, assuming them to be true, is the subsequent derivation from them of metrical and tonal principles valid? I shall argue for the view that, while Constraints 1 to 4 are based mostly on uncontroversial claims about cognition, the derivation from them of tonality and metricality depends upon a number of hidden assumptions and unjustified leaps of reasoning, and upon the *a priori* exclusion of a number of other modes of cognition.

I turn now to a more detailed consideration of Constraint 1. Lerdahl begins 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 may be, it inhibits the inference of structure.\(^{66}\)

This “presupposition” might look like a plausible claim about cognition, but in fact it leaves out a number of quite perceptible attributes that a signal may have:

1. **Continuous change in one or more parameters.** Lerdahl gives no reason why gradients should not play a structural role in music. In fact, they do, and our minds are well set up to deal with them, for good reason: things around us get faster, nearer, louder, not suddenly but gradually. Thus, a ‘presupposition of the theory’ excludes without justification a structural principle which is at work in much atonal-ametric music, and which uses a cognitive facility which we all have.

   Let us take Ligeti’s *Atmosphères* (1961) as an example of the style Lerdahl is presumably alluding to (for he does not mention a specific work). If one listens to the opening of this work (Example 1) one can hear immediately that the simple idea of one event following another is inadequate for music of this kind. Perhaps bars 1 to 22 should be regarded as a single event. If this is so, it becomes clear that the *interior* of each event in this music is as important as the sequence of events. Lerdahl’s point of view, which assumes the existence of a level of atomistic events (notes or chords) whose interiors are structurally irrelevant, is wholly inadequate for this piece. The first event in *Atmosphères* is full of change in several parameters: timbre, pitch region, intensity, and harmonic density; sometimes changes in different parameters coincide, sometimes they intersect. I see no reason why such attributes of the musical surface should not be considered as a kind of structure.

   It is interesting to note also that even in cases of continuous change, where the musical signal is not divided into discrete events, it does not follow from this that the musical surface is incapable of being articulated\(^{67}\) by the listener, for it is quite possible for the listener to impose an articulation that is not present in the signal. A clear example of this is the sudden changes that are a well-known perceptual effect of minimalist music, even though the change in the signal is gradual. Thus, the mind imposes an articulation that is not present in the signal. Given this, it is apparent that very little music would fail Constraint 1 completely.

---

66 Lerdahl, 1988, p. 239.

67 I avoid Lerdahl’s grammatically-loaded term ‘parsed’, since it has not been established that a grammatical metaphor is appropriate. At this point the use of grammatical terminology appears as one more subtle, yet question-begging, means of setting up tonality as the yard-stick against which potentially radically different kinds of musical structure are measured.
(2) Blurred onsets of events. Lerdahl gives no reason why the “blurring” of distinctions between events should be problematic. The smearing of event-onsets does not make the event imperceptible, or even difficult to detect, any more than clouds are imperceptible because they lack a definite outline. Take for example the event in bars 21 to 29 in Example 1: this event is characterised by an accumulation of tremolandi—an independent event if anything is, but one which has no clear beginning; the initial pitch oscillations seem like a detail within the texture of the previous event, and the fact that it is something rather different becomes apparent only after the first few bars (23-25).

(3) Simultaneous events. Events may be superposed, overlap each other, or one event may interrupt another. None of these render the events imperceptible. For example, the entry of the wind and violins at bar 13 in Example 1 might be regarded as another event overlaid on the event that was already happening.

It is apparent that a certain degree of arbitrariness is involved in applying the notion of ‘event’ to music such as this, in which there is often no clear distinction between a gradual change and a new event. Of course, there are clear examples of both—but there are also cases which could be thought of either way: is the crescendo in the horns at bar 19 an independent event, or a transformation of the pitch focus and timbral quality of a larger event?

Research into music perception is still regrettably note-based, so there is little experimental data about the way in which absolute textural or spectral change is perceived and remembered. However, it seems reasonable to suppose that the cognition of such music is closely related to the perception of change in our extra-musical sonic environment. In principle there are indefinitely many ways of dividing up the acoustic signal that we receive in “ordinary life”: for instance, when one hears the sound of a horse ridden by, is it one event (horse passes), two events (horse gets nearer, horse gets further away), or should each impact of hoofs be regarded as a separate event? In practice, however, it would not be surprising if it were found that we group together all sounds that can reasonably be attributed to the same source, and simply form a representation that has something like the content ‘a horse was ridden past’. Clearly this is the most useful grouping strategy for non-vocal environmental sounds—it is of obvious evolutionary advantage both to identify the object-event as such (e.g. ‘movement of predator’), and to notice changes in behaviour within the object-event (e.g. ‘getting nearer’).

I would suggest that the case is similar with much post-tonal music. If the cognition of metrically-subdivided phrases makes use (as Lerdahl and Jackendoff have themselves suggested) of mental capacities originally evolved for the apprehension of speech, then it seems likely that perception of music lacking phrasal structure makes use of our other auditory cognitive capacities—those evolved for tracking the behaviour of physical objects.68 Let us call an ‘object-event’ any acoustic phenomenon (either a continuous signal or a series of iterations sufficiently temporally proximate and

---

68 Note that not all atonal music is aphrasal—the atonal music of Schoenberg, for example, generally retains phrasal structures. Similarly, aphrasality does not entirely exclude tonal functioning—though clearly it would be a rarefied form of tonality that does not involve phrases. The connection between phrasality and metricality seems to be somewhat stronger, though obviously metre without phrase is possible; while phrasality may operate with a localised and hierarchically shallow form of metre, as in plainchant.
timbrally similar) that can be attributed to a single source. Furthermore, to the terms ‘atonal’ and ‘ametric’, let us add the term ‘aphrasal’ to refer to music constructed on non-phrasal principles. I propose two categories for investigating the structure of aphrasal music: that of event—which, like the object-event of non-musical cognition, has a meaning more specific than ‘something that happens’; and that of internal activity—everything that is heard to happen within an event.

Denis Smalley has written at length on the role played by analogues of the behaviour of physical objects and environments in electroacoustic music. I would suggest that, while much of Smalley’s work is specific to the electroacoustic domain, this model of listening is also appropriate to atonal-aphrasal music in general. Smalley has coined the term ‘spectromorphology’ to describe the change of spectral profile over the duration of a sound. In instrumental music, the idea of structuring large-scale activity by analogy with the onset, continuation, and decay of sounding bodies has been explicitly adopted in various ways by “spectral” composers such as Grisey and Dufourt. Just as an individual sound-event has an envelope over the course of which its spectral profile changes, so a group of instruments may combine to produce an analogous effect—a composite or “metainstrumental” event. I would suggest that a version of this model of musical activity can be applied to aphrasal music in general—being, as it is, the principal alternative to the phrasal mode of listening, and thus a good candidate for the mostly likely way for a listener to attempt to understand a signal which does not have a phrasal structure.

This is not, of course, to suggest that the listener actually mistakes a group of instruments for a single sounding object, but merely that atonal-aphrasal music involves the same mental capacities that we use to organise non-vocal aural information: the functioning of these abilities in music need not be tied directly to particular sources or source-types, but nonetheless may involve the same general spectral and morphological attributes that are involved in the recognition of environmental sounds. If we can accept that our cognition of tonal-phrasal structure makes use of capacities evolved for speech processing, without phrasal music being mistaken for speech, then we should have no trouble accepting the possibility of other types of auditory cognition operating in a more flexible way as well: whatever the success of the modular approach, the mind is certainly not that rigid. Let us therefore

---

69 See for example Smalley, 1992. Smalley uses the term ‘indicative field’ to describe such perceived attributes as gesture, utterance, behaviour, motion, and “object/substance”. As his terminology suggests, he is concerned mainly with semantic connections between music and human experience, many of which are possible to a much greater extent in the electroacoustic medium than in instrumental or vocal music. My suggestion, however, is that, in atonal-ametric music in general, our capacity to perceive and remember such properties as these enables us to form more abstract, structural relations—comparable in complexity to those of tonal-metric music—in addition to whatever semantic content the sounds may have.

70 This idea bears a resemblance to Smalley’s notion of ‘surrogacy’, which he uses to describe the remoteness of a sound from physical or gestural origins. In electroacoustic music, a first-order surrogate is a sound whose source is readily identifiable, a second-order surrogate is a sound that retains a relatively strong perceptual connection with physical objects and gestures, whereas a remote surrogate is a sound in which ‘neither gesture-type nor source can be surmised’, but in which ‘links with gesture need not be entirely lost’ (Smalley, 1992, p. 524). However, this terminology cannot be carried directly over into my discussion. For Smalley, instrumental gesture remains within the category of first-order surrogacy, whereas I have would suggest that the very (meta-)instrumentality of a sound forces the listener to adopt a stance which bears a superficial resemblance to second-order or remote surrogacy. This difference should not be overstated, however; for while Smalley is primarily considered with the degree to which a listener can construct a potential source, I am concerned with the use of cognitive capacities for source-recognition and behaviour-inference without the inference of an actual source or behaviour being relevant.
extend the notion of object-event to apply to a sound or succession of sounds which, though it does not necessarily cause in the listener the belief that they are the output of a single source with its natural envelope and resonance, nonetheless leads the listener to use the same cognitive capacities as she uses for recognising and interpreting the behaviour of non-vocal sounds in order to group it as if it were an object-event in the (non-musical) environment.

Recent works of Boulez provide a good illustration of this kind of structure, since many of them can be understood in terms of a very familiar kind of object-behaviour: the onset and decay of a struck object. I refer to this as the *impulse-resonance* model. In Example 2, from *Dérive*, I have indicated how such a model might be an appropriate way to understand the perception of this music. In this example the onset or impulse is “smeared” by flourishes of varying lengths.

Lerdahl in fact alludes to music of this kind, as we can see from

*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 .... 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.71

Lerdahl here raises two objections to the kind of music I have been describing: first, that the structure on levels above that of the “event” is hierarchically flat, leading to the “beads on a string” effect; second, that there is little possibility of complex structuring on levels below that of the event, for if there were too much structure on this level of “subevents”, the effect of a single event would be lost. Thus, music of this kind is hierarchically shallow.

In a way reminiscent of the ontological argument, Lerdahl derives a theory about the perception of aphrasal music from the very idea of hierarchy. It must be remembered that Lerdahl is not presenting empirical evidence about the perception of aphrasal music, to the effect that it is structurally impoverished; rather, he is arguing that, given the nature of cognition, of hierarchy, and of events in this kind of music, it *must* be structurally impoverished. Let us examine how this conclusion emerges.

---

Lerdahl affirms the importance of hierarchical structuring in Constraint 2:

*Constraint 2*: The musical surface must be available for hierarchical structuring by the listening grammar.\(^{72}\)

In support of this constraint, Lerdahl points out that most of human cognition relies on hierarchical structuring (...). 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.\(^{73}\)

In incorporating the term ‘listening grammar’, which, as we have seen, has already been used to refer to the tonal system extracted from GTTM, into this constraint, Lerdahl might appear to assume what he sets out to prove. To avoid this problem, and in the light of Lerdahl’s general comments about hierarchy, I shall interpret the constraint as merely claiming that the mind of the listener needs somehow to organise the signal hierarchically, if she is to be able to remember it.

It should be fairly uncontroversial that music completely unavailable for hierarchical structuring is rare indeed. Lerdahl’s point would appear to be not merely that hierarchy is good and enables the listener to remember, but that deeper hierarchies are better and enable the listener to remember more. This is connected to his notion of ‘complexity’, of which deep hierarchical structure is a necessary condition.

Thus, Constraint 3 deals with the importance of grouping boundaries at a level above that of the event:

A distinctive transition corresponds 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. ...

*Constraint 3*: The establishment of local grouping boundaries requires the presence of salient distinctive transitions at the musical surface.\(^{74}\)

Although Babbitt and Reich, according to Lerdahl, fail on this count, most post-tonal music abounds in such ‘changes in a musical dimension’. Indeed, much recent music abounds in them to a far greater extent than most tonal-metric music. As I shall argue later, these absolute parametric shifts assume a much greater structural importance in the absence of tonality, metre, and phrase.

Of course, for a deep hierarchy, groups must be perceived on several levels: this is where Constraint 4 comes in.

*Constraint 4*: Projection of groups, especially at larger levels, depends on symmetry and on the establishment of musical parallelisms.\(^{75}\)

\(^{72}\) Lerdahl, 1988, p. 239.  
\(^{73}\) Lerdahl, 1988, p. 239.  
\(^{74}\) Lerdahl, 1988, pp. 93–240.
Symmetry is the ‘approximately equal division of a larger time span,’ and parallelism is, roughly, the relation that holds between events perceived as similar. Lerdahl points out that ‘in the 1950s and 1960s literal repetition was widely regarded as aesthetically inexcusable.’ We have to ask how the word ‘literal’ got in here—the point about parallelism, which in itself is quite innocuous, is being surreptitiously strengthened. Perception of similarity does not depend on literal repetition: our lives outside music contain very little literal repetition, and our ability to anticipate events based on memories of past events is useful precisely because it is flexible, and is able to pick out the similarity between events that are not simply repetitions of previous events. The introduction of the idea of literal repetition is quite arbitrary.

What, moreover, must be repeated? It is easy to get the impression that Lerdahl is looking for repetition of themes or motives: however, because much post-tonal music is atemonic, we must look for repetition and development on other levels: timbres, pitch regions, textures, morphologies, densities, activity-types—in fact, kinds of similarity more closely related to those we perceive in non-musical contexts. Our capacities for such perception are much subtler than they might at first appear—consider our ability to distinguish, without even thinking about it, a closing door from a footstep or a breaking twig, the breaking of a twig from that of a branch; and not only this but also the kind of room or environment that in which an event takes place, and how far away it happens. Although to my knowledge perception and memory of analogous musical features have not been empirically tested, the obvious extra-musical utility of source-recognition and identification of object behaviour in non-musical contexts suggests that such cognitive capacities might play a important role in atonal-aphrasal music, and thus that the scope for hierarchical structuring based on parallelism is greater than Lerdahl seems willing to acknowledge. I would suggest therefore that Lerdahl is giving a narrow, tonal interpretation to Constraints 2, 3, and 4.

The possibility of hierarchical structures, distinctive transitions, and parallelisms based on non-phrasal principles is illustrated in Example 2. Here, events on the lowest level are mostly composed of a grace-note figure followed by a sustained, trilled, or tremolando chord. This sometimes gives the effect of a smeared onset (to use our object-event analogue), although in bar 6 we see a more sudden attack. The envelope shapes for each event are shown for the lower levels of event—these envelopes may perhaps be regarded as the aphrasal counterpart of the lowest levels of a time-span tree. Of course, fragments of trees might have been used, as in the example in parentheses—here, the arabesque figure is regarded as a kind of anacrusis to the following chord. I would resist this approach for aphrasal music, however, since it retains the notion of one event being subordinate to another. But this seems inappropriate, at least to this lowest level of events: is the onset of an event subordinate to the decay, or vice versa? Even if we were to answer this question, it is clear that we would be dealing with quite a different notion of subordination to that used in tonal time-span reductions. Significantly, this notion of eventhood does not

75 Lerdahl, 1988, p. 240.
76 Lerdahl, 1988, p. 240.
77 Lerdahl, 1988, p. 240.
depend on tonal stability conditions, and thus provides a counter-example to Lerdahl’s argument from the principle of hierarchy to the necessity of tonality.

The nesting of these event-morphologies can be seen in bars 4 and 5, to a depth of three levels (h, f, and d, with each level being composed of an onset/arabesque and a sustained harmonic field). Higher level groups can be formed on the basis of spectral or morphological affinity, parallelism, or by temporal proximity; or, as in metrical structure, a new higher-level event may be indicated by a stronger low-level event. On the level marked ‘c’ one can see the role of parallelism, with the piano initiating each event on that level. Level b is more open to question. I hear the first four events on level c as a group, giving rise to a more static passage during bar 6. The highest level, however, whose first event ends at bar 27, is undeniable (Example 3).

No definitive quality is claimed for my description of this passage, and it must be acknowledged that the absence of phrasal structure permits a greater degree of ambiguity of segmentation than is generally the case with phrasal music. My point is simply that local hierarchical structuring can exist unproblematically in atonal-aphrasal music. While the structure that I have indicated may be open to question (in the absence of psychological studies, I can claim only that it is an accurate account of how I hear it), it would surely be implausible to suggest that nothing of the kind is heard, that it is perceived merely as a series of events ‘like beads on a string’. So long as such structures can be inferred, it does not seem to be an objection that conflicting segmentations are possible. Indeed, it might be argued that the presentation of a (relatively) unambiguous totality in any work of art tends to ring false or to appear naïve in the (post-)modern world: in this respect, Lerdahl’s claim that the musical structures that he is prescribing ‘need not be nostalgic’ seems mistaken in a very general sense.

In Lerdahl’s view, much of the interest in tonal music is attributed to the prolongational reduction, which takes account of the tension and relaxation, the directionality, of tonal music, and which operates in “structural counterpoint” to the time-span reduction. It may be objected to my view that, while I have suggested a counterpart to time-span segmentation and reduction, there is no further dimension which might replace the prolongational reduction. However, in the discussion of the example from Ligeti’s Atmosphères, I pointed out that a division into events is insufficient to account for everything of structural significance—that if we adopt the object-event model, every event has a significant interior. Our time-span segmentation accounts for the events, but has little to say about their behaviour—the internal articulation or trajectory, their spectromorphology. The envelope shapes that I used to indicate onset-regions convey something of the dynamic morphology of each onset, but the “resonance” of each onset in the decay segment of each onset-decay sequence can clearly have a very rich interior.79

78 Lerdahl, 1988, p. 236.

79 Of course, the envelope-shapes used to indicate onsets might have been extended to show the dynamic morphology of the subsequent “resonance”, but since this would not convey the other properties of the interior activity, I have restricted the use of envelope shapes to indicate event-onsets only, and whether they be abrupt or “smeared”, and so on. 

The inclusion of a level on which a segmentation is made between the “onset” and “resonance” parts of the event might also be called into question. However, since these two parts of an event are clearly perceptually distinguishable while at the same time being grouped together, it seems reasonable to count them as a level of segmentation for the purpose of comparing the depth of perceptible hierarchy with that of time-span segmentations of tonal-metric music.
I do not wish to propose internal spectromorphology as an exact analogue of tonal prolongation; however, I do think that the interior of events can have properties of tension, relaxation, and directedness. These properties are quite different from their tonal-phrasal counterpart, however, since they do not rely on tonal functionality, but rather on properties such as absolute pitch, harmonicity, density of activity, harmonic distribution, timbre, and dynamic morphology. A simple example is provided by the flute in bar 2 of Example 3: the flute crescendo is directed towards the onset of the next event on levels c and d (this is paralleled by the violin in bar 4). Internal rearticulation, absolute pitch, and harmonic quality also increases the tension of an event: thus, internal tension increases during the opening from the opening event on level d (simple decay), through the subtle internal rearticulation of the flute in the second event on this level, the trill in the third, the tremolando trill in the violin in the fourth, and the more active interior of the fifth event (which I have also analysed as a superposition of events on a lower level). This tension is released somewhat in the middle of bar 6, with the reduction of internal activity, dynamic, the lower absolute pitch, and the simpler harmony. (It must be remembered that harmonic tension in this context is “absolute”—that is, it depends not on the function of a chord relative to a system, but on absolute harmonicity and distribution in pitch space, which are psychoacoustic, not tonal, properties.)

It is apparent from these considerations that Lerdahl does not consider all the possibilities when he derives tonality and metre as necessary conditions of multiple-level hierarchical structures. As I noted above, the importance of metre is not really argued for: it seems to be lifted from GTTM without justification. The first mention of metre in “Cognitive Constraints” is the following.

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.80

Lerdahl seems to rely here on the intuition that metre just is hierarchy in the temporal domain. But we have seen that multiple hierarchical levels of events are entirely possible in the absence of metre.81 In order to justify the logical step from hierarchy to metre, we need the additional property of periodicity at all levels. This is the “symmetry” of Constraint 4, which seems to me the one requirement in these first four constraints that ought to be rejected as a necessary property of the musical surface. The closest Lerdahl comes to providing some support for the inclusion of symmetry, and thus periodicity, as an obligatory property is the following passage:

Most music is metrical, and the immediate location of events is established largely in relation to this or that strong or weak beat. As inability to assign a metrical grid

---

80 Lerdahl, 1988, p. 240.
81 Metre is certainly a sufficient condition for hierarchical structure, but Lerdahl would need it to be a necessary condition in order to argue in this way.
weakens the precision of location of events, resulting in a quality of suspended rhythm.\textsuperscript{82}

It is true enough that a very small change to the durations in an ametrical series of impulses is less readily noticed than a similar change to a metrical passage (especially if the change is introduced at a point where \textit{rubato} would be inappropriate). What does this show? Very little, I should think—unless one believes in an absolute aesthetically relevant minimal duration change, specified perhaps in milliseconds. But there is no reason to believe in such a thing. Perceiving changes in \textit{absolute} duration is a different kind of thing to noticing displacements relative to a metrical grid: there is no reason to compare them. Talk of the ‘precision of location of events’ merely disguises the pointlessness of such a comparison. At any rate, Lerdahl underestimates our perceptual abilities if he is claiming that we need a constant and explicit grid of strong and weak beats in order to remember when an event occurred; one might as well insist that all visual art be done on graph paper, in order to facilitate more accurate representations of the location of objects.

With respect to pitch hierarchy, Lerdahl writes:

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)…\textsuperscript{83}

This move is entirely illegitimate. Lerdahl fails to distinguish between two quite different types of hierarchy: the event-hierarchy of the musical surface and the systematic hierarchy of tonality. It is true that events in music are very often \textit{pitched} events, but the term ‘pitch events’ almost seems designed to confuse. Having recognised this, all that one can justifiably assert is that since “events” in music are usually pitched, event hierarchies are normally hierarchies of pitched events. Lerdahl’s second sentence identifies “pitch hierarchy” with tonality. But tonality is a systematic hierarchy, not an event hierarchy: so, understood in this way, the first sentence is plainly false. But if ‘pitch hierarchy’ in the first sentence means something else, then the second sentence is a \textit{non sequitur}.\textsuperscript{84} What is needed to save this argument is the premise that an event hierarchy of events that are mostly pitched has a systematic hierarchy of pitches as a necessary condition. And to this premise my above examples of event hierarchies in atonal music are counter-examples, as are the hierarchies involved in “most cognition” that Lerdahl himself draws attention to.

Thus, neither the necessity of metre nor that of tonality flow from the importance of hierarchical structure set out in the first four constraints. It might be objected, however, that the kind of hierarchical structure that I have described for atonal-ametric music is shallow compared to that of tonal-metric music. However, there is striking evidence that the higher levels of GTTM-style analyses are not psychologically real at all. In a

\textsuperscript{82} Lerdahl, 1988, p. 241.
\textsuperscript{83} Lerdahl, 1988, p. 242.
\textsuperscript{84} The apparent intended structure of the argument is:

\[
p \supset q \quad q \supset r \quad \therefore r
\]

but it is invalid because \(q\) in the first premise is a different proposition to \(q\) in the second.
study of the importance of tonal closure over various time spans, Nicholas Cook\textsuperscript{85} found that for all but the very shortest extract (well under one minute), subjects exhibited no statistically significant preference for the original versions of pieces by Classical and Romantic composers over modified versions in which the tonal organisation was altered so that the piece or extract did not return to the home key; nor did they rate the modified versions as less “complete”, “coherent”, or “expressive”. Cook concludes that these tests ‘indicate that tonal closure has psychological reality for the listener only when the time-scale involved is very small—much smaller than is the case in most tonal compositions.’\textsuperscript{86} The perceived depth of tonal hierarchy would appear to be somewhat less impressive that Lerdahl’s assertions might lead us to believe. As Cook writes, these results suggest that the tonal unity of a sonata is of a conceptual rather than perceptual nature, in contrast to the directly perceptible unity of a single phrase. And it follows from this that the theories of Schenker, Meyer, and Lerdahl and Jackendoff are better seen as a means of understanding the practice of tonal composers than as a means of predicting the effects of their compositions upon listeners.\textsuperscript{87}

In other words, theories involving deeply nested hierarchies are useful, but the analyses that they give rise to should not be mistaken for a description of the heard structure. Ironically, the conclusion that such structures are conceptual rather than perceptual is reminiscent of Lerdahl’s own criticism of serial structure.

This calls into question Lerdahl’s notion of complexity, which is his primary criterion of aesthetic value. In a recent paper Lerdahl contrasts this notion with that of “complicatedness”.

The criterion of value I advocate in \textit{Cognitive Constraints}... has to do not with reaching the public but with a distinction between “complicatedness” and “complexity”. Complicatedness is a function of the number and variety of events per unit time at the musical surface. Complexity is a function of the structural depth that the listener is able to infer from a surface. ... Many 20th-century composers, sensing the lack of complexity in their music no matter how ornate their poietic methods, have sought recompense in complicated surfaces.\textsuperscript{88}

Lerdahl appears to be suggesting that the deeper the hierarchical structure revealed by a GTTM-style analysis, the more “complex” the music. But we have just noted the inappropriateness of identifying too closely the structure revealed by GTTM \textit{qua} analytic method with the kind of structure that might plausibly be attributed to musical experience. In fact, Lerdahl and Jackendoff acknowledge that they are describing an idealisation, when they write:

Instead of describing the listener’s real-time mental processes, we will be concerned only with the final state of his understanding. In our view it would be fruitless to theorise

\textsuperscript{85} Cook, 1987.
\textsuperscript{86} Cook, 1990, p. 55.
\textsuperscript{87} Cook, 1987, p. 204.
\textsuperscript{88} Lerdahl, 1997, p. 425.
about mental processing before understanding the organisation to which the processing leads.89

However, if the high-level tonal organisation does indeed, as Cook suggests, become perceptually irrelevant quite suddenly when the time-span reaches beyond one minute (or less), then one might be led to question even the theoretical significance of Lerdahl and Jackendoff’s idealised listener, in whose mind is represented the “final state” structure described by their theory—for this is an ideal that would never even come close to being realised in actual listeners. The theory thus stands in marked contrast to Chomsky’s description of sentence structure, which applies to structures which almost certainly have a psychological reality for actual speakers.90 But whatever the status of this idealisation, one must be aware that the structural depth ‘that the listener is able to infer’ is by no means to be identified with the structural depth the analyst is able to reveal by applying the rules of GTTM. One cannot simply take a tree diagram for a tonal piece and saying, ‘look how much more complex this is than an atonal piece.’

Thus, not only is perceived hierarchical structure in atonal-ametric music much deeper than Lerdahl acknowledges, but the perceived hierarchy in tonal music is markedly shallower. Thus, while differences of depth may yet exist, Lerdahl dramatically overstates them in support of his aesthetic conclusions.

Furthermore, it seems reasonable to suggest that, even if we were to ignore the higher levels, we would still not have a realistic measure of the complexity of the musical experience. Not being the ideal listener, I do not end up with a tree diagram of even a small section of the music “represented” in my mind—even if we restrict ourselves to trees for the lower levels only, they would appear to be both too complete to be plausible candidates for what one remembers, and yet at the same time to fail to capture many aspects of musical experience which could justly be thought to contribute to “felt complexity”. After hearing a piece, I remember the last things I heard, some themes, various fragments, and at best little more than the broad outline of the form.91 As Eric Clarke writes,

[i]f pieces of music are mental constructs derived from a physical signal, then presumably that signal is usually acoustical, and unfolds over time, rather than[n] being entirely co-present as it is in a score. The sorts of generative rule that could perform such derivations must operate over (and in) time, and cannot be the kind of abstract timeless statement that characterises the rules proposed by Lerdahl and Jackendoff.92

This is almost, but not quite, the point I wish to make. It is not the case that there is a necessary difference between what is derived from a signal presented in time, and what can be described in a synchronic representation. It is not logically impossible that

90 I skirt around any issues about the nature of psychological reality. Suffice it to say that in the current context it can be understood in behavioural terms as requiring at least that the subject be able to repeat or paraphrase the sentence, to form beliefs on the basis of it, and of course to have no difficulty understanding the end of a sentence in terms of the beginning (for example, relating an adjective at the end to a noun at the beginning).
91 Swain (Swain 1986) suggests an upper limit on memory of three or four events per hierarchical level. One is also reminded of Schoenberg’s comment, reported in Rosen, 1972, p. 35: “Where are my favourite passages?, Schoenberg is said to have exclaimed on seeing Schenker’s diagram of the Eroica; ‘Ah, there they are in those tiny notes.’”
92 Clarke, 1986, p. 16 (with correction).
the “final state” of my understanding of what I remember of the piece can be (at least in part) described by GTTM.\(^3\) What I do want to say is that the experience must be quite different to the structures revealed by the rules of GTTM: GTTM does not account for perceived complexity, because it emphasises an abstraction of psychologically improbable depth above several other components of musical complexity which are aspects of musical experience but not of the final-state representation, most notably anticipation. Anticipation in atonal-ametric music is, as I remarked above, closely connected with the interior of events—and thus with the small-scale activity that Lerdahl attempts to trivialise by categorising it as mere “complicatedness”. This reveals a fundamental misunderstanding of the functioning of surface detail, which, as I have argued, participates in parallelism, directionality, and segmentational ambiguity—hardly trivial categories—even when it is not itself subject to hierarchical segmentation.

Perhaps Lerdahl is led down this path by a misconception about the role of hierarchy in cognition. It is true that we tend to represent the world hierarchically, and a signal that does not permit such organisation is difficult to understand and remember. But our hierarchical organisation of the world is \textit{not} primarily a way of remembering as much as possible on many levels, but of selecting an optimal level for the formation of associations. To take a very basic example: remembering that there is a chair, a bed, and a desk is for most purposes more useful than the level above (remembering that there is a room of furniture, for instance) or the level below (remembering that there are lots of pieces of wood). As we progress outwards from the “optimal band” of representations, we get representations that are less and less useful: remembering the shapes of each surface, or the individual flowers on the floral-print upholstery. Hierarchical structuring of stimuli is as much a way of forgetting as of remembering: it is \textit{not} a way of remembering detail on many different levels, but of \textit{not} remembering detail on unnecessary levels. To suggest that remembering in detail on many levels at once is a fuller use of “the potential of our cognitive resources”\(^4\) is misguided, since this is not what our cognitive resources are set up to do. Hierarchical structuring is a means of focusing attention, not of dispersing it.

But for Lerdahl, hierarchy becomes an end in itself. It is one thing to claim, on the basis of evidence about the nature of human cognition, that the imposition of hierarchical structure is important to remembering whatever is structurally important in the musical surface (where these properties are not necessarily themselves hierarchical); it is quite another to say that it is the hierarchical structure \textit{itself} that it is important to remember. (Recall Lerdahl and Jackendoff’s claim that GTTM describes the “final state” understanding of the music.) It is perhaps this principled abstractness of the structures that Lerdahl emphasises that leads James Boros to write:

To my mind, were Lerdahl’s two “grammars” to exist, and were they perfectly to coincide, and were listening experiences to be strictly limited to the generation of hierarchically arranged mental structures, there would be absolutely no reason to

---

3. This skirts around the issue of the temporal order of events affecting the “final state” understanding of the piece by causing one valid interpretation to be preferred over another in a passage that is ambiguous according to the grammar. (See Clarke, 1986, pp. 10–11). While I do not think that a “grammar” should take this into account, it does show that the understanding of a grammar as explaining a “final state” understanding is an oversimplification.

perform or listen to music. Instead, we could content ourselves with burst-like, robotic exchanges of perfect, complete brain wave patterns .... 95

If I am correct, then this impression results from the move from hierarchy as means to remembering the non-hierarchical, to hierarchy as an end in itself. Thus Lerdahl’s reply,

Come now, surely he knows that all human experience, musical and otherwise, depends on brain activity for its existence. 96

is wide of the mark (though invited, perhaps, by Boros’s science-fiction presentation of his point). It is not the means by which the representation is formed that is at issue, but the abstraction of the representation that is the “heard structure”. 97 In this respect, Lerdahl’s insistence upon “cognitive transparency” 98 in music is telling—to insist that one be able to “see through” the music to the structure it embodies is to reject the relevance of using both hierarchical and associative capacities to follow the development and interaction of (“opaque”) sounds—but Lerdahl argues as if it followed naturally from precisely this ability. I am not advocating a clear distinction between the structure and content or substance of music. My point is rather that if Lerdahl wants to base his claim upon the general importance of hierarchical structure, then he must acknowledge that in most cases the hierarchical structuring of a signal is a tool for noticing things that are not themselves hierarchical structures—we do not generally aim to “see through” our everyday experience to the abstract structure that it embodies.

Since atonal-ametric music operates on such different principles to tonal-metric music, it should come as no surprise that different things need to be cognized in order to understand such music. The idea of cognizing music, however, makes no sense as an absolute—all we can succeed or fail to do is cognize something about the music. Thus, we need to specify what it is important to cognize, before we can determine whether some kinds of music are harder to cognize.

What does Lerdahl have in mind when he says that listeners are unable to assign a ‘precise mental representation’ to Le Marteau? 99 Or when he comments that the ‘inference of structure’ is inhibited in such music? Contrary to what some writers seem to imply, we cannot look inside people’s heads and see their mental representations. As Dennett points out, intentional states are attributed to people on the basis of patterns of behaviour, and it is certainly not obviously the case that there must be, in

95 Boros, 1995, p. 546.
97 Lerdahl, 1988, p. 234.
98 Lerdahl, 1988, pp. 237, 255.
addition, an internal correlate of these patterns: there is no reason to think, for example, that the structure involved in Beethoven’s brain when conceiving the form of a sonata movement has something in common with the structure in my brain when I hear it.\textsuperscript{100} We need to ask, then, what behaviour are we looking for when it comes to musical cognition.

Two types of behaviour, it seems to me, exhaust the empirical content of Lerdahl’s talk about cognition and representation. The first, to which Lerdahl gives most attention, is memory. The second, to which Lerdahl gives scant attention, is expectation. This emphasis is in line with the focus in GTTM, already mentioned, upon “final state” mental representations rather than real-time experience. Memory may be understood, roughly, as the persistence of a representation (or more specifically, a belief), in a form available to consciousness. The emphasis on memory would appear to serve Lerdahl’s purposes well—is it not obvious that remembering the music is desirable? If the absence of certain features in the music were found to hinder substantially our ability to remember it, we would surely have to admit that these features are desirable, and that music without those features has something wrong with it.

But what are we supposed to be able to remember? Lerdahl does not consider this question, saying simply, for instance, that in 	extit{Le Marteau} the “details of the piece are difficult to learn.”\textsuperscript{101} But remembering musical detail is not like recording the piece on a mental tape-recorder—we need to specify what 	extit{it is} that we remember about the musical surface.\textsuperscript{102} It is clear that it is not necessary to remember absolutely everything about the piece—for example, the amount of vibrato applied to a particular passage, whether the clarinets are on top or in the middle of a certain chord, and so on, are things that we generally consider less important to remember for most tonal music. It would not be unreasonable to suggest that at least some of the most important things to remember when listening to tonal music are those described by GTTM. These concern harmonic, melodic, and metrical structures, but it would be misleading to say that remembering the GTTM-relevant features of the music is the same as “remembering the music”,\textsuperscript{103} without being careful to point out that what is being remembered is information that is specifically tonal and metrical: the degree of the scale and the metrical position of its attack-point. And indeed, these features are easier to remember in tonal-metrical music, and quite impossible to remember in atonal-ametric music, since in the latter they do not exist.

But relevant remembering about tonal music does not involve remembering absolute pitch region, absolute duration, predominant timbre and articulation, variation in density or texture, or rate of activity, or the rate of change of these parameters—yet

\textsuperscript{100} See Dennett, 1987, \textit{passim}.

\textsuperscript{101} Lerdahl, 1988, p. 232.

\textsuperscript{102} That memory of sound is not like a mental tape-recording is the aural analogue of Dennett’s general point about knowledge as being in a sense more like propositions than like pictures. This view will be discussed more fully in section 12. However, for my present purpose it will suffice to avoid the most naive form of the tape-recorder model of memory by acknowledging that some features about the musical surface are remembered, and others are not.

\textsuperscript{103} Cf Lerdahl, 1996:

Nor does [Boros] ask the relevant question with respect to the argument in “Cognitive Constraints...,” namely whether listeners can learn and remember nonhierarchical music as well as they can hierarchical music. (p. 245)
these are all perfectly memorable, and all contribute to the kind of atonal-aphrasal hierarchical structure that I described above. Of course, it is not the case that these qualities are irrelevant to tonal-metric music: this would be to say that orchestration, or the octave in which a part is scored, or the way a note is attacked, are all irrelevant. It is just that they need not be remembered in order to make sense of the work on a larger scale—they are of essential importance, and may make a lasting impression, but are remembered only fleetingly. They are aspects of expression, not of “structure”.

I think we can say the same about relative pitch in atonal music: while the absolute pitch region of an event may play a structural role, the intervals that make up that event are “expressive” in the same sense that articulation is expressive in tonal-metric music. The intervallic content of an event is essential to the quality of that event—some pitches will be preferable to others, just as some articulation in tonal music will be preferable to others. In other words, while I have no intention of making a fully-fledged theory out of this, I think it can be said that in aphrasal music the “expressive” and “structural” dimensions are reversed. The “relative” features assume a local, expressive function, while in the articulation of larger structures or trajectories, “absolute” differences are most significant. Whereas in tonal music structure is articulated according to the paradigm of metrical and phrasal structure, by cadences, or on higher levels by new themes or changes of key, it is more common in aphrasal music for structural transitions to be marked by “absolute” differences—a change in rate of activity, in predominant timbre, in absolute pitch region. Alternatively, long-term trajectories in a piece may involve gradual change in these parameters, which is not really possible in terms of tonal closure or of modulation, and which would tend to threaten the identity of a theme.

Conversely, in atonal music of this kind, “relative” listening is more relevant within the regions of activity that are characterised by a given set of “absolute” characteristics. For instance, within a section dominated by a certain rate of activity and pitch region, it is possible to set up a pitch centricity through the emphasis of a certain pitch or set of pitches. Within this region, individual pitches may be heard as centred around this pitch or set of pitches. Similarly, quasi-phrasal shapes may emerge from the internal activity of an event. Thus, on the sub-event level, the focus is shifted from the absolute to the relative. But because this is only local, and does not play a role in larger schemes, these features—these intervals, for example—do not need to be remembered for a long time, but only long enough to get us to the next moment, or long enough to impart an expressivity to the local activity. This does not make it unimportant, any more than articulation is unimportant in tonal music. It simply means that to insist on abiding detailed mental representations is to miss the point. Furthermore, for this reason, it does not matter if the pitch domain for the entire piece is large, because scales do not operate over the entire piece: a smaller set can be given local predominance, permitting the recognition of pitches within a given region of activity, without the definition of a scale applicable on higher levels.

I could make an analogous point about metre and duration. Structurally important change in atonal music is often articulated through absolute changes in rate of activity, or in predominant duration. It is true that on this level small differences are less noticeable than they would be if there were a metrical grid against which to judge
durations. But on a smaller scale it is easy to notice small differences—say, between a quintuplet and a sextuplet. And it is not uncommon for a sense of pulse or even metre to be set up locally. Of course, the listener does not need to compare that sextuplet to the septuplet that comes several minutes later, because, once again, these relative differences are of local rather than long-term structural importance. For this reason it is not a problem that there are more than 7\(\pm 2\) durations in the entire piece.

Finally, I think a similar thing is possible with harmony—that is, harmony can be quasi-functional locally without playing a role in long-term structure. This can be heard in the music of composers such as Berio, in which it is possible, sometimes unavoidable, to hear certain progressions as, for instance, moving from dominant to tonic, even though the pieces are not organised according to functional tonality on higher levels.

Thus, in atonal-aphrasal music, properties like position in relation to a pulse or quasi-metre, scale degree, distance from pitch centre, and even tonal function, are relevant on a local level as expressive materials. This allows for the comprehension of detail, without necessitating the memory of this detail over long periods of time, because comparing this detail to temporally distant detail is not musically significant. Meanwhile, higher levels of organisation are articulated mostly in terms of absolute differences, and it is these that must be remembered in order to make sense of the structure of a piece. Now surely it cannot be said that these absolute differences do not match the structure of the mind. In fact, the normal cognition of auditory data bears an obvious similarity to cognition of atonal-aphrasal music, and a marked dissimilarity to tonal-phrasal music. In non-musical cognition we rely on absolute acoustical properties, such as absolute pitch, spectral content, texture, intensity, envelope shape, and so on, to form long-term relationships (such as attributing two temporally distant signals to a similar source), whereas what I have called “relative” listening applies more often to the short term—to interpret vocal inflections, for example, and to infer the behaviour of an object. We do not live in a march or a waltz (or in C major)—this kind of hierarchy is certainly not the kind of structure that ‘most human cognition relies on’.\(^{104}\)

It is not so much the absolute absence of pitch-centricity and periodicity or even metricality, but their optionality, that characterises atonal-ametric music. Thus, as well as the various relative and absolute parameters that I have discussed, and those of gradients in absolute parameters, not to mention the expressive potential of stylistic signifiers from other musics, there emerges also the possibility of musical structures based on “meta-parameters”: the degree of pitch-centredness, the degree of striation or

\(^{104}\) Cf. Clarke, 1987:

The absence of a hierarchical structure of regular integer-related time-spans forces the listener to base judgements of segmental length on durational properties. The internal representation of formal musical units thus takes on the real-time properties found in more general, non-musical contexts. (p. 232)

This weakening of the hierarchic principle in the time domain leaves a variety of associative structuring possibilities available, which offer a different compositional potential from that of hierarchical structure. Only when the primary compositional aim is to achieve the sort of total organic unity represented, for instance, in highly developed sonata form structures, is hierarchic organization (in both the compositional process and perceptual representation) a necessary requirement. (p. 236)

An analogous point may be made for tonal hierarchy, which is absent to an even greater degree in non-musical contexts.
metricality, themselves become musical parameters in a way which was not possible in
tonal-metrical music.

Memory does not last indefinitely, except in the final-state mind of GTTM’s ideal
listener. Given this fact, we might be led to allow the possibility that, in some recent
music, not only might that which it is good to remember be different, but there might
also be a difference in how long it is good to remember various details for. Even in
tonal-metric music, it is far from clear that a permanent mental representation of a
piece would be something to aspire to, were it possible. For one thing, there would be
no reason to listen to a piece more than once. (Indeed, pieces which are very easy to
remember, because of their simplicity or repetitiveness, are pieces we generally do not
want to listen to many times.) Forgetting, then, can be a virtue. Of course, remembering
some things is essential to perceiving a piece as a meaningful sequence of events
(though some composers have sought to abandon even this).

Lerdahl’s argument seems to involve another non sequitur analogous to that
discussed earlier in relation to hierarchy. Not only does Lerdahl assume that if
hierarchy is necessary for cognition, then a very deep hierarchy must be even better; but
also that, if remembering details is good, then remembering them for longer must be
better. Now this is true up to a point if you want the listener to understand a motivic-
thematic “musical argument” in a sonata, and the importance of remembering themes
and motives explains the repetitive nature of classical music. But this is not the case
for a great deal of recent music, which is concerned with trajectories through absolute
acoustical dimensions such as those discussed above. In this music, literal repetition
would be pointless and arbitrary, since it would be facilitating long-term memories of
details that need only be remembered in the short term, and in so doing would
interrupt the evolution of the larger-scale tendencies, rather make them more readily
perceptible. This does not mean that this activity is unimportant: it is remembered,
even if fleetingly; some fragments may be remembered longer, and they can be
recognised as instances of activity-types which the piece may set up as relevant in the
long term, even if the detail of each tokening of that type is only of local relevance.
Thus, detail may be integral to large-scale structure even if it is not remembered, since it
may contribute to a trajectory—indeed, might be the only possible detail at that point
which would have the appropriate effect on the perception of subsequent
events—without being remembered far beyond the perceptual present.

This amounts to an affirmation of the importance of considering not only memory,
but also anticipation. There is no reason to think that remembering is an end in itself, or
that aesthetic pleasure is derived simply from remembering things. It is more likely that
it is derived from anticipating things, and from having one’s expectations confirmed or
disconfirmed to varying degrees. In “everyday life”, one doesn’t want to remember
everything, but only those things relevant to predicting the future. Perhaps, then, it is
not unreasonable to suggest that this principle operates in music also. Indeed, it has long been recognised that the enjoyment of music lies to a large degree in the setting up and interruption of expectation.\textsuperscript{105} A detailed memory of the music should thus perhaps not be regarded as an end in itself, but as a basis upon which expectations can be formed about the continuing course of the music. If this is correct, then what ought to concern us is the relevance of the levels of coding operative at any moment in the piece to the formation of expectations. If they are relevant only to the immediate future, as, for example, relative pitch information is in atonal music, then the best use of our cognitive resources is to forget that information when it is no longer needed—remembering it for longer would be pointless and aesthetically irrelevant. If we really want to consider music in the light of our cognitive resources, we need to recognise that our intelligence consists as much in forgetting the right things as it does in remembering. This sits well with my earlier observations about the role of hierarchy in cognition, to the effect that hierarchical structuring is better understood as a way of focusing attention on an optimal band of levels. Now we might identify the optimal levels as those most relevant to forming expectations: hierarchical structuring is a way of distinguishing what to store in long-term memory from what need only be remembered momentarily.

In the light of my earlier comments about the sense of directedness that can be conveyed by the interior spectromorphology of an event, it seems reasonable to suggest that all features of the musical surface may contribute to this sense of expectation, and in particular to the experience of one sonic characteristic giving rise to another. This aspect of listening is concerned with trajectories on a single level: thus it is not hierarchical, but nor is it simply a question of counting events: thus it would appear to constitute another aspect of felt complexity that is absent from Lerdahl’s account.

None of this is to say, however, that there is no place for enduring memories in atonal-aphrasal music. The effect of a trajectory along an absolute parameter (for example, a texture that becomes increasingly horizontally dense, or rises in pitch in a stochastic manner, or becomes increasingly inharmonic) may have a very clear psychological effect over many minutes. The degree to which this is the case would need to be tested experimentally, but my prediction is that trajectories of this kind, which play a major structural role in much atonal-ametric music, will be found to be noticeable over many minutes. It seems reasonable to suppose that this effect would be due to both the absolute nature of parametric change (it is easier, for instance, to remember an absolute quality, such as inharmonic, or fragmented, or sparse, than to remember a relative quality which exists only in relation to a conventional grid, such as a scale), and to the relative continuity that is possible when such trajectories are articulated (thus not necessitating the representation of a large number of discrete values). The two types of activity I have drawn attention to—generalised trajectories and locally relevant detail, are characteristic of stochastic music (broadly defined).\textsuperscript{106}

\textsuperscript{105} This idea was first elaborated in depth by Meyer (Meyer, 1956; Meyer, 1967)
\textsuperscript{106} Xenakis notes that much of the perceptual effect of serial music is best understood as stochastic: ‘The composers [Stockhausen, Boulez and others] thought they were orthodox serialists but that was only true on paper. In reality they had mass events which they should have listened to in an unbiased manner. On the level of conscious thinking they should have introduced such notions as average density, average duration, colours and so on.’ (Xenakis, in Varga, 1996, p. 54) Lerdahl’s assertion that Xenakis’s music illustrates the gap between compositional system and cognized result, that it ‘divorces method from intuition,’ (Lerdahl, 1988, pp. 235–6) is just false.
If the above account is correct—and if Cook is correct in claiming that the large-scale tonal schemes that form the higher levels of Lerdahl’s hierarchy appear not to be psychologically real anyway—then the higher-level organisation of much atonal-ametric music will prove to be much more perceptually significant, extending to greater durations, than tonal organisation.

What leads Lerdahl to leave out of consideration “absolute” properties as a possible basis for the cognition of musical structure? I have already excluded the easy solution of positing an innate structure which defines in advance the kinds of categories that might be relevant. But Lerdahl’s “Atonal Prolongational Structure” may provide a clue as to another reason why Lerdahl assumes that memory and cognition must involve similar categories for all kinds of music. He writes:

The conventional wisdom, at least in the United States, holds that Schenkerian theory explains diatonic tonal music and pitch-set theory explains atonal music (chromatic tonal music is a source of discomfort). This scenario is implausible from a psychological standpoint if only because it presupposes two entirely different listening mechanisms. We do not hear Elektra and Erwartung in completely different ways. There is a good deal of 20th-century music—Bartók or Messiaen, for instance—that moves smoothly between tonality (broadly speaking) and atonality. In short, the historical development from tonality to atonality (and back) is richly continuous. Theories of tonality and atonality should be comparably linked.

...What is needed is a theory that is general enough to underlie both idioms yet flexible enough to adapt to the ways in which the idioms differ and intermix.107

Lerdahl thus argues from the existence of an historical and perceptual continuum between tonality and atonality to the necessity of a single theory which explains the cognition of both. This does not follow. I can imagine a continuum of objects between roller-skates and steamrollers, but this does not imply that they are usefully understood as being instances of a single kind. Periodicities makes his point seem more convincing than it is by means of a judicious choice of examples: if he had written, ‘we do not hear The Art of Fugue and Pithoprakta in completely different ways,’ it would be somewhat less convincing. It may be more accurate to say that tonal and atonal music have rather different modes of perception, but that aspects of both can coexist in a single piece of music. Erwartung, for instance, retains many features of tonal music: phraseology, expressive signifiers, and a directional harmonic language, for example. Likewise, we might say all of these things are present in Elektra, but that traditional tonal functioning is already being eroded. In fact, with regard to much of Strauss’s music it has to be said that high-level tonal organisation is scarcely more present than in atonal music. (Thus, if Lerdahl is right in claiming that something is amiss in music

---

without tonal hierarchical depth, then this is a problem not only with the avant-garde
music he targets, but also with a considerable amount of late Romantic music, where
tonal functioning is almost reduced to low-level harmonic directionality, which is still
present in the “atonal” music of Schoenberg and Berg).108

These facts more than adequately explain the similarity between “the way we
listen” to these two works, while allowing that the difference between clearly tonal and
clearly atonal music might be so fundamental as to necessitate a different theory for
each. And this view, unlike Lerdahl’s, also accommodates the fact that where there is
very little similarity in other domains, as in The Art of Fugue and Pithoprakta, there is no
interesting sense in which we hear the music in a similar way.

I think this reveals another possible reason why Lerdahl attributes various problems
of contemporary music to psychological universals rather than to cultural phenomena:
he seems to think that, necessarily, either these problems are due to mere unfamiliarity
of the style, or to the structure of the mind, the latter being something that cannot be
changed. But what I am suggesting is that it is not clearly either of these. Through all
the stylistic change from the Renaissance to Stravinsky certain things remained
important—appreciating music involved, among other things, following the melody,
remembering the melody so that you notice when it comes back, following the harmonic
progression, and forming expectations about the next harmony, and all of this was
dependent upon a segmentation based on metrical properties. The expectations about
melody and harmony change dramatically between styles, but until this century it has
gone without saying that remembering the tune was a good thing to do, that dividing
up the music into phrases was important, that harmony was functional. These are all
activities that have become less relevant in much music of the second half of this
century. No degree of familiarity with this music is going to help if one still expects to
hear these things—no amount of exposure will enable the listener to remember many
themes in Xenakis, or the phrase structure of the Ligeti’s Atmosphères: Lerdahl is right
about that. But Lerdahl assumes that these same things must be important (he says so
explicitly in the passage quoted above), and fails even to consider what other things
might be important in music if these are absent. Of course, it is an empirical possibility
that the answer is ‘nothing’—but this cannot be assumed, and in fact I have shown that
it is very far from nothing.

This assumption is symptomatic of the fact that the constraints were derived or
“generalised” from a consideration of tonal music—there is no way for abilities that
are of little relevance to tonal music to make their way into Lerdahl’s picture. Again, a
generalisation of this kind would be appropriate if the aim were to determine the
structure of an innate grammar, but if the domain is cognition in general, then this
approach is unavoidably partial. It is analogous to “generalising” from horses to
animals in general, and then arguing that fish are deficient because they lack legs.

108 It might almost be said that, in the case of the Strauss of the Elecktra period, that the rare appearance
of recognisable tonal progressions has little to do with tonal structure, behaving more like
conventional signifiers. As Cook writes, ‘[i]n such works the traditional tonal ending is being used
not because it is tonal (as Rosen suggests, there may be little or no large-scale tonal organization in
such works), but because it is traditional: in other words, it has become simply a conventional sign of
finality.’ (Cook, 1990, p. 44.)
When Lerdahl applies an adapted version of GTTM to Schoenberg, he relies heavily on categories which this music shares with tonal music. His analysis of the Schoenberg extracts are similar to what they would have been if plausible tonal chords (whose degree of stability did not contradict the salience implied by the other conditions) had been substituted in the passages. Lerdahl’s salience conditions adapt his tonal salience conditions simply by leaving out those conditions relating to tonal stability, and effecting a corresponding increase in emphasis upon the remaining conditions. Thus the analysis says little of interest about atonal music in general, and very little indeed about atonal music which does not have a metrical, dynamic, or phrasal structure evocative of a quasi-tonal form of event salience and closure.

I have been working with the notion of two kinds of listening, which I have called ‘relative’ and ‘absolute’. Admittedly, these terms are far from ideal. What I call absolute listening is not absolute in any profound sense: it simply refers to the kind of listening which recognises sounds according to their psycho-acoustic properties, their spectromorphology, rather than their position in relation to a metre, phrase, or tonic. It does not mean, of course, that relationships between sounds are irrelevant; merely that the relationship between a sound and a conventional system is irrelevant. Thus, an interval is a relationship between two pitches, but in an atonal context this interval may function as an “absolute” property, as I have been using the term, if it is the distance between the pitches as such (and of course, the absolute pitch of the pitches) that is important, and not the fact that one is, say, the third degree of the scale and the other the sixth.

Despite their capacity to mislead, I have preferred these terms to another pair which may appear to express a similar distinction: ‘categorical’ and ‘non-categorical’. Diana Raffman, for example, hypothesises a distinction between two levels of pitch perception: that of chromatic pitch (or “C-pitch”) identification, which is the level at which pitches are heard ‘as instantiating the sequence of pitches specified in the score’, and that of “nuance pitches”, or “N-pitches”, which is the level at which we can discriminate much finer pitch differences than those of the chromatic scale. According to Raffman, the knowledge of N-pitches is ineffable, by which she means that we can gain knowledge of these pitch differences, but can neither report nor remember them.

Before concluding, I shall make a detour though the work of Raffman, for two reasons. First, I wish sharply to distinguish my view from hers; for mine is not a theory.
of ineffability. However, despite this difference, it might appear that Raffman’s work could provide the foundation for an objection to my view: for the set of properties that Raffman draws attention to appear to overlap with the set of properties which I have argued can form a basis for the perception of complex structures in atonal-ametric music. Yet Raffman claims that these properties of sounds are ineffable: they cannot be reported or remembered. If this is true, then clearly they cannot be of structural significance after all. Therefore, my second purpose in discussing Raffman’s claims is to refute her entire theory of ineffability, and thus avoid this objection.

Raffman points out that, on the one hand, there is strong evidence that we are unable reliably to categorise intervals much finer than chromatic semitones, yet, on the other hand, it is clear that we can discriminate much finer pitch differences without categorising them. Raffman’s model seems to be that the physical signal gives rise to a “nuance-level” representation, which in turn gives rise to a “mental score”. It is only the latter that we can conceptualise and remember. However, we still have conscious access to the earlier “nuance” stage of processing, despite our inability to capture its nature in language. As evidence of this, she cites the fact that we can hear details finer than the mental score would allow—for example, out-of-tuneness:

The first thing to realize is that, even if real-time perception of musical intervals is to some extent categorical, the effect is far less pervasive than Burns and Ward and Siegel and Siegel (among others) have imagined. I say this because it is obvious that we just do hear out-of-tune playing all the time, and we hear it in systematic, inter- and intrasubjectively uniform, easily testable ways. We spend lots of time and money getting our instruments and voices properly tuned; musicians lose their jobs for playing out-of-tune!!

The obvious question that arises is: why does Raffman think that this kind of thing is “ineffable”? It is as easy to say ‘a little too sharp’ as it is to say ‘minor third’. Why does she think that there is an important difference? I think the source of this idea lies partly in Raffman’s model of musical cognition, whereby the C-pitches are the effect of a schema being brought to bear on a lower-level pre-categorical representation. But this is not the only possible model. One could instead posit, for example, a “perfect fifth detector” that, given a reference pitch, marks all pitches in a certain range as qualifiers for the label ‘perfect fifth’. In this model there is no inference from a “nuance” representation of pitch that a certain interval should be labelled as, say, ‘perfect fifth’; it is simply the case that pitches in a certain range are picked up by the perfect fifth detector. The only place the “exact” pitch is represented would be the basilar membrane, or thereabouts—at any rate, we would not need to be conscious of it.

I expect that Raffman would object to this model, responding that by making all representations categorical, I am no longer able to explain the facts that her notion of the nuance-level seeks to account for: namely, that we can hear things like out-of-tuneness, vibrato, and so on. But these facts do not imply an ineffable level of representation or knowledge, but merely that, as well as detecting perfect fifths, we can detect pitch fluctuations, the beating of out-of-tune notes, and so on. This does not

---

112 See, e.g., Burns and Ward, 1982, 247; quoted by Raffman, 1993, p. 84.
113 Raffman, 1993, p. 76.
make the representation of properties like vibrato or out-of-tuneness any less
categorical, or any closer to the physical reality of the sound. This should not surprise
us, since it is just as easy to report out-of-tuneness and vibrato as it is to report a given
interval.

The whole case for ineffability rests on the distinction between categorical and
continuously variable perception. But this distinction does not deserve the reverence
normally given it. ‘Slightly higher’ is a category. The only distinction to be drawn is
between (categorical) identification (as in ‘perfect fifth’) and (categorical)
discrimination (as in ‘slightly higher’). Raffman’s case for “ineffability” turns on the
point that we can discriminate pitch differences much smaller than we can categorise.
But in making this point she is comparing two entirely different abilities—the ability to
recognise an interval, and the ability to say which of two pitches is higher or when a
pitch is out of tune. It does not follow from the difference in “fine-grainedness”
between these two abilities that we have two kinds of knowledge, one effable and one
ineffable; it simply means that we can make various kinds of report about pitch—one
involving intervals to the nearest semitone, another about whether a note is out-of-tune,
and another about whether a given pitch is higher or lower than one just heard.

Of course, this may be merely a disagreement with Raffman’s terminology. But if
this is so, it is difficult to see what is gained by, or what is interesting about, the
practice of referring to judgements about tuning as “ineffable”. Unless, that is, one is
thereby referring to some “experience” beyond the grasp of words: this would be
reminiscent of the view held by philosophers such as Thomas Nagel, who claim that
the experience of qualities such as colours, smells, and sounds—or qualia, as they are
called—constitute facts that can be neither exhaustively expressed in words nor
scientifically accounted for.114 But Raffman explicitly rejects the idea that her theory of
nuances is a form of qualophilism.115 At any rate, if hers were a view akin to Nagel’s,
then “what it is like to hear a perfect fourth” would be just as ineffable as “what it is
like to notice that a note is slightly higher”.116

Raffman briefly considers a related objection in the following passage.

It may be objected that whereas you know how the acoustic events of the
performance sound qua C-pitches, you precisely do not know how they sound qua
nuances; and that that difference makes all the difference where claims of knowledge
are concerned. This complaint smudges an important distinction, however; for although
it is true that you don’t know how the events of the performance sound qua tokens of
particular nuance types, you certainly do know how they sound qua nuances as opposed
to, say, qua C-pitches. In other words, you hear them as nuances, and therein know that
they are nuances (as opposed to, say, C-pitches); it’s just that, for lack of internal
schemas, you don’t know which nuances they are.117

The problem with this is that the knowledge ‘that they are nuances’ is not ineffable or
unreportable at all. This is why, I think, we see Raffman here resorting to talking about

114 See, for example, Nagel, 1974.
115 Raffman, 1993, pp. 126ff.
116 Raffman comes closer to the “what-it-is-like” style of ineffability in her “type-2” or “feeling”
ineffability (see Raffman, 1993, pp. 37–62), which she distinguishes from nuance ineffability, which is
type 3.
‘how they sound’—in other words, despite explicitly rejecting a phenomenological interpretation of her position, she nonetheless slips back into affirming the existence of qualia. This comes out clearly in her consideration of another objection—that knowledge is by nature verbally expressible and that therefore nuances cannot be knowledge. She responds:

Don’t I know the determinate pitch of a tone when I am listening to it? or the determinate shade of an object when I am looking at it? 118

The insistence that ‘surely I know my own experience!’ is a well-known intuition-pump from the qualophile literature. The big problem here is the notion of ‘determinate pitch’. If ‘knowing the determinate pitch of the tone I am listening to’ is just an odd way of saying that I am listening to that pitch, then it is true; but nothing is gained by referring to all experience as ‘knowledge’. If, however, knowing the determinate pitch amounts to noticing that it is, say, 538.5 Hz, then no, we do not know the determinate pitch of the note we are listening to.

Raffman insists, however, that it makes no difference to her view whether we agree with her that such ineffability be called ‘knowledge’:

Although I find it plausible to regard the N-level as an instance of musical knowledge, … the success of my project does not depend on this; in other words, the success of my project does not require that the N-level be an instance of knowledge. 119

Few distinctions turn out to be merely terminological, and this is no exception. If it is not to be described as ‘knowledge’, then this will be because it shares too little with the other things we call ‘knowledge’. In order to avoid vacuity, Raffman needs to specify something which makes this N-level more than simply a fact of experience; otherwise the content of Raffman’s claim vanishes into the truism that we can only experience something while we are experiencing it. 120

Raffman’s view also seems to have the consequence that we ineffably “know” all of our experience—every sensation on our skin, every shade we see, every moment of ambient sound. That is a lot of knowledge to attribute to ourselves without good reason. It is implausible to suppose that this much information is actually represented to consciousness as a continuous stream of data about the world. 121

118 Raffman, 1993, p. 90.
120 It is unclear for the most part whether Raffman has in mind a kind of knowledge by acquaintance (connaître) or a special kind of knowledge by description (savoir) which somehow lacks a verbal expression for \( p \) in the formula ‘A knows that \( p \)’. (At the very least, knowledge by acquaintance seems to be what she has in mind with ineffability 2: see Raffman, 1993, pp. 37–62.) But whatever the N-level is, to call it ‘knowledge’ is to misuse the word. I would suggest that we cannot make much sense of the verb ‘to know’ if it is used in such a way that it operates in an entirely different way from verbs such as ‘to believe’, ‘to expect’, ‘to remember’, ‘to desire’, ‘to recognise’—how can you have knowledge by acquaintance admits of the possibility of remembering and recognising (and probably expecting and desiring).
121 This notion of the “Cartesian Theatre” of consciousness is one of Daniel Dennett’s favourite targets. As Dennett remarks, referring to his example of a room decorated with Marilyn Monroe wallpaper, ‘The hundreds of Marilyns in the wallpaper seem to be present in your experience, seem to be in your mind, not just on the wall. But since, as we know, your gaze can shift in a fraction of a second to draw information from any part of your visual environment, why should your brain bother importing all those Marilyns in the first place? Why not just let the world store them, at no cost, until they’re needed?’ (Dennett, 1993, pp. 359–60) So, if it’s not remembered, chances are it wasn’t in your brain in the first place!
The view I have been espousing has much in common with a view of consciousness held by philosophers such as Dennett and Rorty, both of whom reject the notion of unreportable knowledge. Raffman attacks what she describes as Dennett’s ‘propositional’ account of knowledge by claiming that it cannot account for nuance representations—thus, supposedly, avoiding the qualophile phenomenology that is Dennett’s traditional target. Raffman proposes

to show that the nuance ineffability of musical performances suggests an independent line of attack against Dennett—a line that does not, at least not on the face of it, require the endorsement of a phenomenology and its problematic accessories. Dennett’s theory leaves something out all right, but what it leaves out is the conscious representational (as opposed to phenomenological) content of the N-level. 122

This would be all very well if it were possible to make sense of Raffman’s assurance that we are, in some sense, conscious of these within-category nuances—that is, if it could be demonstrated that we are capable of something more subtle than making judgements such as ‘a bit too much vibrato’ or ‘a little sharp’, despite our inability to attest to them. But given that we do not have any behavioural evidence of these representations—and, it would seem, no way of getting any—what evidence can we possibly have of them at all, other than phenomenological evidence of the ‘I really do have them’ variety?

In fact, Raffman seems to misinterpret Dennett in a quite general way. She quotes the following passage in Dennett’s “Quining Qualia”.

Pluck the bass or low E string open and listen carefully to the sound… Now pluck the open string again and carefully bring a finger down lightly over the octave fret to create a high ‘harmonic’. Suddenly a new sound is heard: ‘purer’ somehow and of course an octave higher…But then on a third open plucking one can hear, with surprising distinctness, the harmonic overtone that was isolated in the second plucking. The homogeneity and ineffability of the first experience is gone, replaced by a duality as…clearly describable as that of any chord…There is nothing to stop further refinement of one’s capacity to describe this heretofore ineffable complexity. 123

But Raffman paraphrases this passage as follows:

Be the details as they may, any such speechlessness is merely “practical”: given sufficient time and language mastery, you could provide an exhaustive report of the contents of your conscious perception of the guitar tone…

Indeed, not only will you be able to report that you are hearing an E-natural guitar tone; the implication seems to be that, eventually, you will even be able to report that you are hearing a tone of a certain determinate pitch, timbre, loudness, and so forth. 124

I think this interpretation of Dennett is mistaken. Dennett is not saying that we could report everything about the note, but simply that what are unable to report is not really an aspect of the content of the “experience” at all, even if we think it is. In Dennett’s view, we are simply mistaken about many of the things we think we

123 Dennett, 1988, pp. 71–2
experience. It might be, for example, that in most cases there is nothing like a “determinate pitch” of our experience. Dennett is not suggesting that, given enough time and the requisite vocabulary, we could say everything that makes a guitar sound a guitar sound; indeed, this would be an absurd position—even having provided verbally a Fourier analysis of the sound, we would still have to describe the experience of the sine tones, and so on, *ad infinitum*. No; rather, Dennett is suggesting that in many cases there may be no content to our experience that is not captured by, say, ‘low guitar note’. If there is more, it is describable and effable—but there need not be more. Of course, this is not equivalent to the claim that hearing a guitar sound and being told about it are the same experience—obviously they are not. The claim is simply that we are mistaken if we think that there is some kind of fact that can only be known subjectively through introspection of that experience.

Thus, when, in formulating her objection, Raffman asks, ‘[h]ow will [Dennett’s view] treat our conscious awareness of the particular determinate N-pitch of that E-natural guitar tone?’ we can nip the objection in the bud simply by responding that we do not have awareness of anything like a ‘particular determinate N-pitch’. There is no reason to think that we have anything of the kind. Certainly, experiments show what the limit is on our perception of, for instance, the difference in pitch between two notes heard in succession. But the content of the judgement made by subjects in such experiments is simply ‘yes, it still seems slightly higher’, or something like that—nothing ineffable. The fact that we can tell the difference between, say, 30 slightly different E-naturals does not mean that we have a separate “content of experience” for each one—it simply means that we can say, up to a certain limit, whether a certain pitch is higher, or a certain interval larger, than a certain other pitch or interval.

Dennett’s guitar example, once again, leads Raffman back to a phenomenological argument:

Suppose you wanted to convey to another listener the determinate N-pitch of that guitar tone (or, if you prefer, the determinate shade of some object); in other words, you wanted to convey to him how the E-natural(23) sounds. ... How would you accomplish this? So far as I can see, your only option would be to show him, to present him with an instance of the E-natural(23); there would be no telling him how it sounds. What I want to suggest is that an analogous challenge confronts your auditory processor in its effort to “convey” to conscious awareness—to convey to you—how that E-natural(23) sounds. So far as I can see—and I don’t know a less metaphorical or misleading way to say this—your processor must somehow present the E-natural(23) to conscious awareness.126

I suspect that Raffman here does not appreciate the full force of Dennett’s position. Dennett insists that there is *no such thing* as “how the E-natural sounds”. Thus there is nothing that needs to be explained by a “presentation to conscious awareness” of the kind Raffman envisages. The heart of Dennett’s position is that it is philosophically and scientifically pointless to insist on a difference that does not make a difference. For Dennett, to attribute a belief to someone is to make a judgement based on that person’s behaviour, verbal or otherwise, and at the same time to make a prediction about that person’s subsequent behaviour. Raffman’s nuances cannot be remembered, recognised,

---

126 Raffman, 1993, pp. 140–41.
or identified—and thus it is difficult to see what behavioural pattern we could be taking as the basis for an attribution of knowledge. The discrimination tests cited by Raffman cannot possibly play this role, because such data necessarily consists of entirely effable, verbalisable judgements which have no bearing on the existence of an “N-level”.

Raffman, if I read her correctly, considers an objection similar to the one I have just made, when she writes that

one might suppose that your conscious awareness of the E-natural(23) could consist in a cluster of comparative judgments discriminating it from other N-pitch standards. The comparisons would have to be performed in imagination, as it were; for if they required the actual sounding of standard tones, then the content of your conscious awareness of the E-natural(23) would not be fully determinate until all of the relevant standards had been sounded and comparisons made; that is to say, the content of your episode of awareness of the E-natural(23) would not be determinate at the time of its occurrence.\(^{127}\)

The idea considered here—that there is nothing to one’s “conscious awareness” of the E-natural(23) but a collection of comparative judgments—is along similar lines to my own view. However, Raffman’s reply does not affect my position, aimed as it is at providing a *reductio ad absurdum* of the idea that there is nothing to the awareness except a cluster of judgements. For I am perfectly happy to accept the supposed *absurdum*: that there is no such thing as an “episode of awareness” with a determinate content of ‘E-natural(23)’. If there is any reason to believe in such a thing, Raffman certainly has not provided it—and I cannot see what could purport to be a reason, short of the phenomenological introspection that Raffman has, rightly in my view, already rejected. In short, Raffman cannot have provided a ‘representation-based, as opposed to qualia-based’\(^ {128}\) objection to Dennett’s view, for Dennett is also a verificationist, and there can be no verification of N-level awareness that is not inherently qualophilic.

Both Raffman’s idea of nuances and my notion of absolute listening are set up in contrast to the categorisation of pitches in relation to a scale. But Raffman’s dichotomy is misleading, for it groups together the knowledge of continuously variable (yet entirely memorable) properties, with spurious cognitions of unnameable attributes, relegating both to the domain of “ineffable knowledge”. Raffman and Lerdahl share the idea that attributes of the performance which do not form part of the “mental score” (or “heard structure” in Lerdahl’s terminology—in both cases something like a GTTM-style reduction\(^ {129}\)) are difficult or impossible to remember. While for Raffman, in contrast to Lerdahl, these cognitions seem to be positively valorised, this is at the price of denying them any structural significance, and consigning them to the role of an experiential supplement to structure. In effect, Raffman’s view implies that the psychoacoustic properties that contribute to the perception of structure in tonal music (and which form the basis of the “mental score”), on the one hand, and those properties which contribute to “expression” (small pitch differences, vibrato, and presumably other

---

\(^{127}\) Raffman, 1993, p. 141.
\(^{129}\) Raffman, 1993, pp. 49–53, 64–73.
properties if the theory were extended beyond the domain of pitch\textsuperscript{130}), on the other, have the function (structural or expressive) that they do as a result of some kind of psychological necessity. If correct, this theory would pose a threat to the view I have argued for, namely that a kind of role-reversal takes place in much post-tonal music—scalic, metric, and pitch-centric properties taking on a local “expressive” function, and absolute differences assuming a structural role. However, I have argued that unstriated, continuously variable, “absolute” properties are neither unreportable nor unmemorable, and are quite able to form the basis for complex musical structures; and indeed that the whole notion of ineffable unmemorable knowledge is incoherent.

\[\text{xiii}\]

We have seen how Lerdahl falls into the trap of making both hierarchy and memory ends in themselves, which leads him to lose sight of their actual function in normal cognition—despite the fact that, having ruled out the modular interpretation, Lerdahl’s constraints must be interpreted as being derived from our general cognitive abilities and limitations. Ultimately, Lerdahl’s narrow interpretation of the role of hierarchy and memory in cognition lead him to an unconvincing conception of musical complexity, which he upholds as a criterion of aesthetic value. I argued above that this fails to take account of the music as experienced in time, and in particular of the role of anticipation: this is not necessarily a problem for an analytic method, but it is a big problem for an aesthetic stance.

Yet it seems to me that Lerdahl’s conception of musical complexity would remain impoverished, even were he to incorporate into it the idea of anticipation. For, while semantic and connotative associations are not my main concern here, it must be pointed out that such associative aspects of the musical surface, which are possible to at least as great a degree in atonal-ametric as in tonal-metric music, can no doubt make significant contributions to perceived complexity. Nor can semantics be ruled out as irrelevant to considerations of structural complexity and cognition; for an event that bears some meaning for the listener is more likely to be remembered, and thus will have an impact on the perception of structure: not only in that a repetition will be more likely to be perceived, but also in that an additional domain of possible relationships is opened up: events can be heard to be related because of a semantic relationship which does not consist in a structural similarity. This is perhaps most obvious in works which carry traces of and references to other styles, yet of course is not limited to the explicitly quotational.

\textsuperscript{130} Even remaining within the domain of pitch, Raffman’s view seems to necessitate assigning various other properties to the realm of the ineffable. For example, while Raffman does not address this question, it seems unavoidable that Raffman’s view (that the cognition of pitch fluctuations significantly smaller than a chromatic semit is ineffable) be extended to much larger differences which are also not categorised in terms of the chromatic scale. For example, the pitch difference between a low and high cymbal, or between two widely separated clusters, would also appear to be only ineffably known according to Raffman’s view. This point further highlights the discrepancy between Raffman’s view and my own.
Thus, to my earlier point that complexity in atonal-ametric music involves not only the hierarchical structure of events, but also the surface detail of those events, we can add the recognition that complexity involves not only parallelism in the form of isomorphism, but also semantic relations. And of course, the boundaries of relevant remembering are not clear-cut: there is room in musical experience for the not-quite remembered, and for uncertainty about whether you’ve heard something before. Surely these also contribute to the complexity of music experience.

If there is one strand that runs through all of my objections, it is the suspicion that, once Lerdahl’s view is clarified to the point where we can recognise its empirical content, once the status of the purported universals is worked out, and once the science is separated from the scientism, all Lerdahl succeeds in demonstrating is that atonal-ametric music has little in the way of tonal and metrical structure. This is not a very interesting result, and I have urged that we should not be too worried that this is the case, for the non-obligatory nature (not necessarily the absence) of tonal and metrical structures in atonal music allows perceptually complex structures which are not possible in a tonal-metric context.

This is, of course, not to deny that something was lost when tonality became impossible for many composers. Perhaps most significantly, music lost a rich harmonic implicational structure, which allows a listener tacitly to judge the relative probability of harmonic direction on several levels. A similar point might be made about metre and phrase. This is no small loss. But I would resist the idea, initially tempting, that what has been lost is musical grammar itself—that there has been a transition from the grammatical to the “agrammatical”. If true, it would certainly provide an easy response to Lerdahl: if he is presenting constraints on musical grammars, then they cannot apply to music which has no grammar.

However, I think that this would be to interpret too literally the idea of musical grammar, which is more of a metaphor than is often admitted. Lerdahl and Jackendoff’s theory, despite its apparently Chomskian inspiration, bears only a superficial resemblance to generative grammar. For one thing, it is not generative in Chomsky’s sense. Despite Lerdahl and Jackendoff’s disclaimers and qualifications, it remains that their theory generates only analyses, not well-formed strings. It must also be remembered—as indeed Lerdahl and Jackendoff point out—that in music there are no deep structures, rules of transformation, or parts of speech. However, the difference goes deeper than that. Linguistic grammar is not concerned primarily with word order, but with relationships between grammatical categories. Of course, word order in many languages is an indicator of function, but this is only of accidental importance—in Latin, for instance, the words can be in almost any order, and only the emphasis of the sentence, not the proposition it expresses, is altered by an alteration of word order. What is of necessary importance in language is the synchronic relationship between elements in the sentence, not their order. Yet in the case of music, the order is of necessary, not accidental, importance: it could not be otherwise, for there is no musical equivalent of the elements of a sentence (such as noun phrase, verb phrase, and so on)—there are no labels on the nodes of a prolongational tree. Grammar is essentially synchronic, and music, despite Lerdahl’s best efforts to convince us

---

131 Lerdahl and Jackendoff, 1983, pp. 5–6.
otherwise, is essentially diachronic. The existence of grammatical categories in language is necessarily connected with the existence of semantics—even if it be a minimal semantics of the kind found, for instance, in propositional calculus. In other words, that a word is a noun is simply a highly general aspect of its meaning: the line between grammar and semantics in language is only a matter of convenience. Yet the conception of musical structure as involving grammar without semantics would appear to require a principled distinction of the kind that does not exist.

It would seem from these considerations that the idea of musical grammar is a very loose metaphor: while language is only accidentally about word order and essentially about categories ultimately founded on semantics, music has neither language-like semantics nor grammatical categories, and is essentially about the way in which one thing leads to another—and in so far as this is true, it does not point to a further difference between tonal and atonal music on the level of grammar.

I will not dwell on this point here. However loose or tight the analogy between tonal music and grammatical structure, I have shown that the idea that atonal or ametrical music is inherently less rich, or cognitively impoverished, is false—or, if true, then at least it cannot be deduced from cognitive psychology. After all, surely we ought to be suspicious of aesthetic conclusions based on supposed psychological universals, if those conclusions actually prohibit or devalue kinds of music that people actually listen to and claim to enjoy. If Chomsky came across an apparently fully-functional human language that did not fit his innate grammar, he would have to change his theory. Yet we have plenty of music that does not conform to Lerdahl’s grammar: what, then, are people who claim to appreciate such music—who claim to find it as interesting as tonal-metrical music—actually doing? Either they are deluding themselves, or they are lying, or they have non-human brains. None of these answers seems entirely satisfactory. But if we do not like any of these answers, then we must admit that it is a matter of exposure and acquired understanding after all—in which case we are certainly a far cry from innate psychological universals.

132 As Putnam writes:

These linguists [who maintain that there is a sharp line between grammatical and semantical deviancy] would for example reject the category “animate noun” as a permissible grammatical category, although they would of course admit it as a potential semantical category. Now, suppose a Frenchman says, “George gave me a table and I saw at once that she was red and had four legs.” Chomsky would say that this sentence is ungrammatical because the pronoun “she” does not agree with the inanimate noun “table”. These linguists, on the other hand, would maintain that the sentence in question is grammatical, simply because they have no basis for calling it ungrammatical. One wonders how they would deal with such languages as German and French, where questions of gender have long been regarded as grammatical questions: Would they proceed the same way, or would they have one policy for English and another for French, and if so, on what basis? Their position, as I gather it, is that it is only features that are arbitrary, that have nothing to do with meaning, that are properly called grammatical. However, I agree with Jakobson and with Boas, that there do not appear to be any arbitrary features in a language in the sense indicated. For example, if one says that it is arbitrary that we say “She is here,” and not “Is she here,” the obvious answer is that while it may be arbitrary in some absolute sense, in the context of English it is not arbitrary; we use one when we want to make a statement and the other when we want to ask a question. Notice that if we agree that the categories “abstract” and “concrete” should be prohibited in grammar, then on exactly the same grounds we should prohibit “masculine” and “feminine” on the one hand, and “indicative” and “interrogative” on the other. (Putnam, 1974, p. 90.)

Note that the point here is not that grammar cannot be studied independently of semantics, once the line between them has been decided. Rather, it is that what we call grammar may be understood as simply the domain of meaning which exhibits the most useful regularities. But if, as in GTTM, the existence of a language-like semantics is rejected in principle, then it is difficult to see how a properly grammatical domain could be delineated.
I think this conclusion has a more general significance beyond the question of the truth or falsity of Lerdahl’s theory. It shows how easily vague language and tacit assumptions can be brought into the service of conservativism and aesthetic authoritarianism. It points to the misguided nature of attempts to turn the question of the dissemination of post-tonal music from an aesthetic, political, and indeed economic issue into a cognitive-scientific one. In this age when words like ‘accessibility’ and ‘communication’ are used too frequently and with too little understanding, it is of some significance that at least one major attempt to give scientific respectability to the conservative side of the debate fails. On the other hand, this very attempt has forced me to outline the ways in which atonal, ametric, aphrasal music is in fact based not merely on the negation of tonality, metre, and phrase; but on different, yet equally interesting, human cognitive abilities. Thus we have reason to believe in the validity and worthiness of attempts to educate audiences and to familiarise them with new music. The “empiricism” of figures such as Boulez—a long-term advocate of education and exposure—has proved to be somewhat nearer the mark than the rigid and restrictive musical mind of the nativist position.
Bibliography


