4. The Inner Work of Music: Lerdahl and Jackendoff's 'Generative Theory'[1]

4.1 The Inner Work of Music

A Generative Theory of Tonal Music (GTTM, 1983), the collaborative work of music theorist Fred Lerdahl and linguist Ray Jackendoff, is conceived as a *theory of music*, not 'music theory'. While all music theory studies the way music is organized, most studies are based on the assumption that this organization may be described or analyzed as a property of music as an autonomous artefact. The way meaning, emotion and understanding come about in listener's minds have usually been considered not the subject of music theory, but of music psychology and aesthetics. A different track had been chosen by Leonard Meyer, who in *Emotion and Meaning in Music* (1956) introduced elements of gestalt psychology to what was still basically music theory. Lerdahl and Jackendoff however "take the goal of a theory of music to be a *formal description of the musical intuitions of a listener who is experienced in a musical idiom*" (p. 1). Being psychological in nature, the theory is subject to empirical testability: decisive is how well it describes this listener's intuitions.

The thoroughness and explicitness of its method are unprecedented in music theory – still the field of reference. That it does not nearly attain the degree of formality aimed at is, maybe, not surprising. More perplexing is that there is very little in this theory which is actually *psychological*. The authors' approach of the mental is similar to that of Generative Grammar: it aims at providing a grammar which *could* model the computations in the listener's/speaker's mind, without having access to those mental processes themselves. As a *generative* theory (or grammar), it relates by hypothesis to the listener's musical competence, or those mechanisms which allow her to understand music, rather than to her performance, or actual listening behaviour. But since the terms of analysis are almost exclusively those traditional in music theory, it ultimately remains just that: music theory, with a bonus.

The justification for this assimilation of music psychology with music theory can be found in the transformation of a psychological thesis into an ontological one:

> The present study will justify the view that a piece of music is a mentally constructed entity, of which scores and performances are partial representations by which the piece is transmitted. (GTTM p. 2).

The 'work', then, is in the mind, as a final state of understanding, conceived through a process in time, but idealized in the theory as a timeless totality (p. 4).[2] The choice of words is notable. Though neither scores nor performances can exhaust the 'work' concept, a performance (if not a score) is, in common parlance, a 'realization' (even if partial and non-definitive) rather than a 'representation'. The representation is, rather, an ephemeral actuality in the listener's mind. Presenting GTTM as a justification of this piece of ontology, the authors turn things around: it is this view which justifies GTTM. Its implausibility is responsible for most of the problems in the theory.

1 This is a chapter from Muns (2014), in a slightly revised version (2015).
2 E.F. Clarke (1986: 10) speaks of their "'God's-eye' view of musical structure".

The mentalist mode of music's existence implies an analogy to Chomsky's 'I-language' (Chapter 2), but with an important twist. I-language (or 'I-music') is *competence*. A musical work however is not 'music', and the inner reality of musical competence does not imply the inner work of music. I-language is I-grammar, not I-speech. We refer to the score or a performance as a source to feed our mental image; we suppose it is 'out there', and our understanding is a continuous engagement with it. The 'inner work of music' is a Schenkerian-idealist creation, rather than a cognitive object. Schenker approached the musical work as an extra-temporal entity, of which the generative essence could ultimately be defined as the triad. In his reductive vision of musical structure, temporality disappears through the assimilation and equalization of long range and short range types of relations, and an indifference towards the temporal dimensions of form.[1]

Parallel to innate Universal Grammar, the authors postulate a universal musical grammar which is ultimately genetically defined. This is however merely a side-track in their discussion of western tonal music, which is discussed as a grammar in its own right. The authors call their theory 'generative' with explicit reference to the linguistic concept (borrowed from mathematics), meaning the 'generation' of grammatical sentences, that is, the implicit definition of possible sentences by rules. 'Generation' does not mean 'production', and it is the grammar, not the speaker that 'generates'; the speaker however uses the generative capacity of her grammar to produce sentences, as the listener uses the same to parse and understand them. GTTM does not have this symmetry: it is a listener's grammar only, and therefore not actually a generative grammar in the linguistic sense. Katz and Pesetsky therefore call GTTM a 'generative parser'.[2] Lerdahl and Jackendoff attempt to minimize the difference by a shift of interest from grammaticality to structure, and by speaking of sentences as *given*:

> [...] what is really of interest in a generative grammar is the structure it assigns to sentences, not which strings of words are or are not grammatical sentences. (GTTM p. 6, cf. 112)

On this premise (an "unquestioned assumption of actual research in linguistics"), they criticize earlier attempts to construct musical grammars, such as Sundberg and Lindblom's grammar of Swedish nursery tunes (1976), which are in fact productive algorithms for highly restricted musical domains. The criticism is unfounded: such an algorithm is not, as they suggest, merely a decision procedure for grammaticality; it does exactly what linguistic grammar does: it allows the speaker to speak, and does this by defining the set of possible sentences. The grammar, of course, does not talk and in that sense is not 'generative'. The only way of testing its accuracy however is by using it in the creation of sentences or melodies.

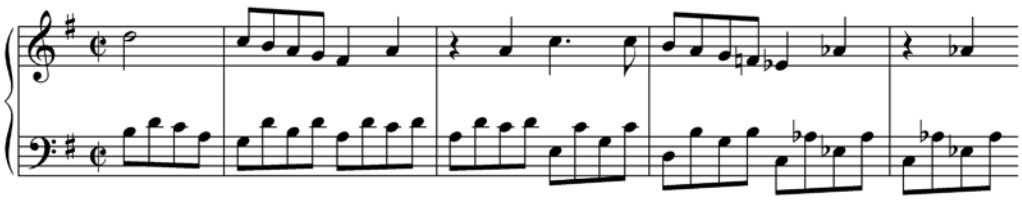
That their focus is on parsing, rather than creative use, is motivated not so much by the correct interpretation of 'generativity' as by the inapplicability of traditional Generative Grammar to music of greater complexity. As they explain in the 1996 preface, this is due to a supposed inherent underdetermination of musical grammaticality: "any grammar we could write generated too many 'grammatical' structures that did not make musical sense" (p. xiv). In other words, in the forest of

1 Cf. Muns (2008).
2 Katz and Pesetsky (2011: 5).

possible parsing trees no plausible, intuitively acceptable tree can be selected on the basis of grammatical rules alone: the Well-Formedness Rules (WFRs) of GTTM cannot exhaustively define the idiom. This does not point towards a distinctive feature of musical grammar, as the authors seem to believe; it is a consequence of the fact that they choose to study *pieces of music* rather than musical grammar. They found it "less rewarding to specify structural descriptions for normative but dull examples than for works of lasting interest" (p. 8). Using the same argument, a linguist might prefer to study poetry instead of trivial everyday sentences; but it would be a bad start for the study of grammar. Linguistic grammar generates sentence schemas, not discourse; for this, contextual input is needed. The authors are aware, of course, of the difference between poetry and prose, and profess their dedication to the "musically mundane" (p. 7); but they fail to draw consequences for the domain of grammar. It is highly likely, after all, that *some* aspects of music are grammatical, or rather, syntactic (harmony being the obvious candidate), while others are supervenient upon the grammar, or indirectly derived from it. In the same way, linguistic discourse cannot be explained by the rules of syntax alone.

The decision to make grammaticality dependent upon intuitions about 'musical sense' is a consequence of this focus upon the work of music. It robs us of the possibility to make some useful distinctions. The formal notion of grammaticality and the vague one of musical sense cannot simply be assumed to fall together or be mutually dependent. Music may be grammatical but devoid of sense, or as Roger Scruton illustrates, "syntactically quite in order, even though bland and meaningless".[1] On the other hand, it is easy enough to produce structures which are clearly ungrammatical, yet have some 'sense' by virtue of certain standard 'lexical' expressions (in the manner of *'Furiously sleep ideas green colorless'*, cf. 2.3; Ex. 4.1).

Ex. 4.1



'Ungrammatical' pieces are much less common in the repertoire than 'bland and meaningless' ones which are grammatical. They are mainly found among those by learners. For instance, a childhood minuet by Mozart (written at age 5, K1a) may be called 'ungrammatical'[2] because of its ineffective and contrapuntally irregular connection of clichés.

Instead of exploiting these distinctions (which suggest a significant linguistic parallel), the authors have chosen to maximize the domain of grammar by relaxing its formality. They therefore supplement their Well-Formedness Rules (WFRs) with Preference Rules (PRs), which indicate

1 Scruton (1997: 180).
2 Gjerdingen (2007: 335).

tendencies to hear in specific ways (described as producing the most 'stable' structures). These are not given any specific weighting; their interaction therefore remains in an undecided state (p. 9, 55).[1]

The postulate of the inner work of music is unnecessary for the construction of a musical grammar, but it does allow the authors to practice psychology with the methods of music theory. In this way they approach their goal through a number of shortcuts. One of these is the reliance on score excerpts as a substitute for music, often speaking of 'notes' as perceptual entities. According to Lerdahl, this was a piece of methodology borrowed from linguistics.[2] The assumption that scores relate to music as written texts relate to language is however obviously false: scores are basically instructions waiting to be realized; written language has in the course of history developed into an independent medium.

The second shortcut from music theory to psychology is the uncritical adoption of an undefined body of traditional theoretical concepts.[3] They accept as "given":

> [...] the classical Western tonal pitch system – the major-minor scale system, the traditional classifications of consonance and dissonance, the triadic harmonic system with its roots and inversions, the circle-of-fifths system, and the principles of good voice-leading. Though all of these principles could and should be formalised, they are largely idiom-specific, and are well understood informally within the traditional disciplines of harmony and counterpoint. Nothing will be lost if we conveniently consider them to be an input to the theory of reductions. (GTTM p. 117; cf. 192, 296)

If the aim is to explicate commonplace basic musical understanding, such commonplace, basic but informal and non-psychological notions cannot simply be taken as data; they should be the first to be rigorously defined. It is rather as if a linguist would simplify his job by referring to school grammar for definitions of 'verb' and 'predicate'. Music theory as such is not directly concerned with perception, and its concepts may be applied to perception only indirectly, *if* these concepts themselves are sufficiently established on their own ground – which they are not, considering such a common, but dubious and misunderstood notion as that of harmonic 'function'. Especially curious is the failure to provide a theory of harmony of any kind; harmonic concepts are simply injected ad hoc into various reductions. In this, Schenkerian influence may play a part. Though Schenker did formulate a theory of harmony, its coordination with the reductive praxis of *Der freie Satz* is notoriously left to subjective (and highly idiosyncratic) intuitions.

The third shortcut is taking personal intuition as a stand-in for experimentally validated judgments, a practice impossible to reconcile with what purports to be an empirical psychological theory (p. 248-49). This is common in generative linguistics; for judgments about grammaticality,

1 After GTTM, Optimality Theory (Prince and Smolensky 1993), originally formulated for phonology, has adopted preference rules (involving potentially conflicting constraints as well as the means to resolve conflicts). Temperley (2004) builds on GTTM, providing quantified PRs.

2 Lerdahl (2009: 188).

3 "It does not seem to occur to them that an elementary obligation on any theory of tonal music (in the sense in which they use that term) is to examine whether the principles they mention are indeed as well understood as they claim [...]." Longuet Higgins (1983: 93). The issue is addressed also in the review by Harvey (1985).

linguists routinely accept their own feeling if it concerns their native language.[1] In some instances at least my intuitions about analytical choices differ from those of the authors.

Contestable is also the authors' broad and a-historic conception of tonal music, which takes no account of stylistic differences between Baroque, Classical and later styles, treating them all as one 'idiom'. Characteristic is their reliance on a Bach chorale setting as paradigmatic for harmonic tonality (4.4 below).

For a more detailed discussion, I will choose four angles of approach: (1) the various hierarchic aspects and their representation by tree diagrams; (2) the Schenker influence and its relation to traditional theoretical concepts; (3) the psychological aspect, and the authors' ideas of what constitutes listeners' intuition; (4) GTTM's contribution to an understanding of the language-like properties of music. The last section of this chapter discusses later proposals based on GTTM.

4.2 Hierarchy

While Universal Grammar is abstracted from the plurality of concrete languages, GTTM intends to define the grammar of one 'idiom', that of western tonal music, even though emphasis is laid on perceptual generalizations which are supposedly universal. For this and other reasons the authors are right in not trying to model a musical grammar on linguistic grammar, apart from borrowing some principles of methodology. This is evident mainly in the use of tree diagrams, which follows from their central claim: that the basic structure of music is, like language, hierarchically organized. According to their 'Reduction Hypothesis',

> The listener attempts to organize all the pitch-events of a piece into a single coherent structure, such that they are heard in a hierarchy of relative importance. (GTTM p. 106)

Perception accordingly proceeds analytically, in "a step-by-step simplification or *reduction* of the piece, where at each step less important events are omitted, leaving the structurally more important events as a sort of skeleton of the piece." The hierarchic concept applies especially to the parameter of pitch, in which all secondary events are heard "in a specified relationship to surrounding more important events" (this they call their 'Strong Reduction Hypothesis').

The kinds of relations permitted and excluded by this formally hierarchic grammar are clearly illustrated by well-formed and ill-formed trees. Excluded are cross relations, multiple branches associated with one event, and omission or incomplete analysis of consecutive events at a given level. Well-formed trees show recursive relations. Where two branches meet, one continues upward; this indicates the dominant event ('head') at that level (Cf. Ex. 4.4 and Fig. 3.5).

Methodological similarity with Generative Grammar hardly goes beyond the basic notion of hierarchic organization associated with rules of grammaticality. In GTTM, the perceived musical hierarchy is one of elaboration, one 'event' (mostly a 'pitch-event') depending upon another which is 'structurally more important' (in the context of GTTM, 'structural' means 'relatively stable' and eo ipso 'higher in the hierarchy'.) In traditional linguistic phrase structure grammar ('X-bar Theory',

1 A controversial issue; see Boden (2008) and references.

largely due to Jackendoff) the tree represents relations among grammatical *categories* (as illustrated in Fig. 3.3 and 3.4). What is at the lowest level a morpheme (or, in other theories, a sound), is subsumed on a higher level under a phrase, at the highest the sentence. Lerdahl and Jackendoff deny that musical structure has such categorial distinctions.

> Rather, the fundamental hierarchical relationship among pitch-events is that of one pitch-event being an *elaboration* of another pitch-event; the latter is the structurally more important event of the two. Thus a suspension is an elaboration of its resolution, the events en route in a phrase are elaborations of either the phrase's structural beginning or its cadence (as the case may be), and so forth.
>
> In these musical cases the event that is elaborated is retained along with the event(s) that elaborate it; the structural beginning and the cadence of a phrase do not disappear or convert into something else in the course of fleshing out the phrase as a whole. (GTTM: 112-3)

'Elaboration' is, just like 'event', a vague abstraction. I find it hard to give any meaning to the assertion that, for instance, a harmonic degree V generally 'elaborates' I just like a suspension 'elaborates' its resolution, or that this kind of reduction represents any listener's perception. In fact, it is just as plausible (if not more) that an adjective or a determiner elaborates a noun and a prepositional phrase a verb phrase – yet they elaborate it in different ways. It is this highly abstract notion of 'elaboration' that allows the dominant event to retain its identity on the next higher level of the hierarchy, instead of being subsumed under another grammatical category (such as 'dominant region' or 'cadence').[1] Leonard Meyer has called this the 'fallacy of hierarchic uniformity'.[2] Arguing for the validity of this kind of reduction, the authors state that the ultimate reduction level – such as an E-flat major triad for the *Eroica* – "is a way of saying what key the piece is in" (p. 109). For this very reason the identity-retention central to their reductive practice does *not* hold up: the triad is merely a symbol for the key, not an actual chord; and the key is not established by any particular tonic in the piece, but by the whole tonal structure. Lerdahl, who calls this uniform kind of hierarchy 'representational', allows that in a representational hierarchy "the specific content of an event degrades and generalizes in memory, depending on the framework in which it is experienced", implying a degree of abstraction on the higher level.[3] This comes close to stating that this event becomes a *symbol* for something different, in the sense that (arbitrarily) a tonic chord may 'stand for' a key. It is difficult to give this notion any experiential (as against analytical) meaning.

Following this "generally pervasive intuition that subordinate events are elaborations of particular dominating events" (p. 116), the authors define hierarchy with the help of set-theoretical concepts as discussed in Chapter 3, as "an exhaustive partitioning into hierarchic, nonoverlapping segments or elements," placing selected elements of the musical organization either within the same set (which constitutes a level) or in a relation of subordination (containment) or superordination (p.

1 A distinction of levels, similar to those of phrase structure grammar, is made in Rohrmeier's *Generative Syntax of Tonal Harmony* (2011), with Riemannian notions of harmonic functions defining phrase levels.

2 Meyer (1967: 96); Meyer (1973: 89); relating to GTTM, Rosner and Meyer (1986). Hierarchic uniformity has been proposed in linguistics (Bare Phrase Structure, cf. 3.6; Katz and Pesetsky 2011: 20), cf. 4.2.1.

3 Lerdahl (2001: 35); 'representational hierarchy': after Cohn and Dempster, *Hierarchical Unity, Plural Unities: Toward a Reconciliation*, in Bergeron and Bohlman (1992: 156-81).

287, 13). Aspects of musical structure which display no hierarchic ordering, such as motivic and thematic relations, are acknowledged but accounted for at best indirectly (p. 9, 17). That these are ignored follows from their conception of grammar as generative. More questionable is that harmony as such is implicitly categorized as non-hierarchic: it receives no separate treatment and is simply included among those features which are "well understood informally". From a generativist perspective this is a contradiction in terms; the formal treatment of tonal-harmonic organization in Lerdahl's *Tonal Pitch Space* (2001) may be seen as an implicit acknowledgment.

Though such hierarchic ordering may in principle provide a transparent account of musical structure, the theory as a whole is far from simple, for the authors discriminate four distinct 'hierarchic dimensions', which are supposedly universal (p. 280), but whose interrelations are unclear: metrical structure, grouping, time-span reduction, and prolongational reduction. Metrical structure is of these the most conventional; grouping structure largely, but not completely conforms to traditional notions of phrase structure. Most problematic are time-span reduction, which depends on meter at lower, and on grouping at higher levels; and prolongational reduction, which in turn relies on time-span reduction while introducing the most explicitly Schenkerian elements and a parameter of 'tension' and 'relaxation' (p. 8-9).

In their choice of these parameters or 'dimensions' there is a marked difference with mainstream theory, which usually takes under consideration meter and rhythm, harmony, melody, and a composite of all of these, at various levels constituting phrases and periods. This is by no means a clean division, since melody and phrase are not independent of meter. That their choice of dimensions is *psychologically* motivated is not obvious.

The discussion of meter proceeds (unlike the pitch-related aspects) from basic definitions. Meter consists of beats, in themselves durationless, which are separated by equal intervals or time-spans. Metrical accent implies that a beat is a beat also at the next higher level (the division hierarchy represented in note values) – and no more; this accent may be marked by stress, but that is not an essential ingredient of meter.

Given time-spans, one may wonder why we need a *durationless beat*. Since a beat is defined by its distance from a preceding or following beat, the notion of a beat without an associated time-span (inter-onset interval or duration) is meaningless. Its origin is maybe the fact that beats/time-spans are usually *marked*, by coinciding with new events; the authors refer to metronome clicks and to the "hypothetical infinitesimal point in the conductor's beat" (p. 18). Watching a conductor, we may have the illusion that he marks such indivisible moments. Instead, we watch a continuous movement which allows us to anticipate the next reversal of direction. This reversal, and the newness of a musical event, may be thought of as point-like; but there is no reason why we should speak of a special timeless 'beginning' of a time-span, and give it the name 'beat'. With these point-like beats the concept of meter becomes unnecessarily abstract and cerebral; it obscures what may well be the most likely origin of our sense of meter or pulse, a sensorimotor capacity for beat entrainment. It is

unlikely that the human species would have evolved a capacity for 'thinking' beats before it *moved* to the beat, and it may be argued that we still keep the beat by interiorised movement.[1]

The more explicit reason why the authors hold on to the paradoxical notion of durationless beats is the fact that metrical beats are associated with a pattern of weak-strong alternations. These, they argue (against the prosodic theory of rhythm of Cooper and Meyer),[2] cannot be carried to levels above the bar. It is not at all clear, however, that meter itself has relevance above the bar level: the metric unit *is* the bar. Meter constitutes the clearest, most straightforward case of hierarchic structure in music, but it is limited in scope; beyond bar-lengths the sense of regular pulsation becomes increasingly hazy. A feeling of regular succession is often sustained by phrases of the same length, i.e. groups of bars, but one may doubt whether this grouping constitutes a genuine instance of meter: precisely because it lacks the clear sense of weak and strong that rules within the bar. In quite large dimensions a sense of longer and shorter may persist and thus may create a less precise but still perceptible *rhythm*. Logically and perceptually, a quantitative sense of longer and shorter may exist without the sense of pulse or beat that defines meter. Meter is often represented as a container of rhythm; it is, rather, a *constrainer*, which adjusts longer and shorter to the proportions of a grid. In their discussion of Mozart's G Minor Symphony (K550), the authors suppose a higher level meter (hypermeter) at work, but have to concede that it is soon subjected to a "metrical shift" (p. 22). By this concession, the operation of hypermeter at this level turns out to be a gratuitous assumption.

4.3 Grouping

Like meter, 'grouping', or the perception of patterns in a sequence of 'events', constitutes a recursive hierarchy, consisting at different scales of motifs, phrases and sections (p. 14-16). Those events are primarily pitches, though what exactly is included in a group is loosely formulated (Grouping WFR 1):

> Any contiguous sequence of pitch-events, drum beats, or the like can constitute a group, and only contiguous sequences can constitute a group. (GTTM p. 37)

In practice, the authors speak of 'notes' at the basic level (Grouping PRs 2, 3), thereby strengthening the close association between structure and its graphic representation.

Traditionally, music is thought to exist of phrases and sentences, rather like language consists of sentences. Factors by which these phrases are recognized are harmonic (especially cadences), melodic contour (including repeated patterns), and rhythmic-metric balance or contrast. Articulation is, as in language, the art of clarifying the phrase, giving it both shape and character, by making the sounds cohere into groups and of separating them, by making them sound prominent or subservient. For this, the composer has to rely strongly on the performer, who has not only to

1 Thaut (2005: 6, 7) argues on the basis of neurological evidence that "the perception of rhythm and formation of rhythm may be biologically based more on the entrainment of oscillatory circuits in the brain than on actual acts of measurement in terms of timekeepers that are often conceptualized and modeled as clocks, pulse counters, or stopwatches in the brain." "[...] the perception of rhythms is not an event-based process, but an interval period-based process, with pulses simply serving as event markers demarcating rhythmic intervals".

2 Cooper and Meyer (1960).

understand the written articulation marks but above all the shape and expression of the melodic-harmonic phrase. The listener, of course, has to be receptive to these nuances; a creative listener may contribute imaginary nuances absent in an insensitive rendering. This is largely a matter of style; and style has strongly evolved since harmonic tonality has come into being.

A generalized approach to 'grouping' is thus complicated by the fact that what is involved are first of all very general features of auditory musical perception; second, general features of harmonic tonality; third, features particular to the stylistic context; fourth, features contributed by the idiosyncrasies of performance. Since the authors are concerned with generalities, in their discussion of grouping they focus on the first of these. This is of course a legitimate choice; but difficulties arise by their attempt to apply such general notions of pattern perception (gestalt principles like temporal proximity and pitch similarity) directly to concrete musical works. Their rules of phrase construction do not consider harmonic progression (harmony being largely restricted to the cadence), and silently pass by theories of musical phrasing:

> [...] the grammar that describes grouping structure seems to consist largely of general conditions for auditory pattern perception that have far broader application than for music alone. Moreover, the rules for grouping seem to be idiom-independent – that is, a listener needs to know relatively little about a musical idiom in order to assign grouping structure to pieces in that idiom. (p. 36)

In the context of classical music this is a wide reductive leap: rather like trying to base linguistic syntax on phonological units, instead of parts of speech. It is with such highly unspecific relations between 'notes' that the GWFRs and GPRs and are concerned, not with the question what kinds of events may be 'grammatically' put in sequence.

Approaching grouping as something more basic than phrase-construction, the authors make grouping independent from meter (p. 26, 280).[1] In the sense that groups or phrases can start and end on any beat or time span, this is obviously correct; beyond that, this view is hard to sustain, unless one limits the 'grouping' concept to highly abstract pitch configurations, which are unlikely to have any perceptual reality. According to the gestalt principles they borrow from visual perception, a pattern of notes separated by rests will be grouped according to the same intuitions as a configuration of visual shapes separated by space. They therefore make the implausible supposition that musical 'events' are perceived, basically, outside a musical, i.e. tonal-metric framework. It is possible that in the absence of such a framework, and confronted with highly indeterminate 'events', we would apply gestalt principles similar or identical to those in visual perception; but as soon as we perceive *music*, we almost instantaneously pick up rhythmic cues which determine the grouping conditions of these events. Grouping PR 3 stipulates that a group boundary tends to be perceived if the notes are relatively far apart, differ dynamically, are differently articulated, or are of different duration (p. 46). In classical music these factors of attack, articulation and dynamics stand in a close relation to meter and metrical accent. The interrelation of grouping and meter is accounted for only

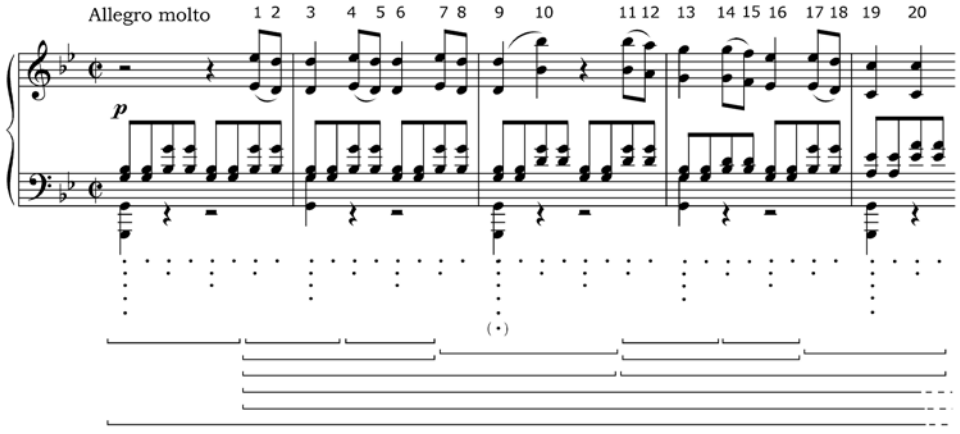
1 Jackendoff and Lerdahl (2006: 37): "[...] rhythmic organization is the product of two independent hierarchical structures, grouping and meter. The relative independence of these structures is indicated by the possibility of dissociating them."

indirectly, in a late stage of the cognitive process, by their Grouping PR 7, which presupposes the high level synthesis of time span and prolongational reductions (p. 52). Meter, however, allows us to anticipate and group events in advance (and therefore allows us to be cheated also).[1] Hearing a non-accented click pattern such as ♪♪ 𝄾 ♪ 𝄾 ♪♪𝄾, we might prefer a grouping by proximity (♩♩ 𝄾 ♪ 𝄾 | ♩♩ 𝄾), but in the appropriate *musical* context we will hear, for instance, 𝄴 ♪| ♪ 𝄾 ♪ 𝄾 | ♩♩ 𝄾. It is not evident or very likely that such 'well-informed' grouping overrides a primary 'raw' gestalt perception.

Accepting that meter is a factor in grouping does not imply that grouping and meter should run parallel at all levels, creating gigantic prosodic feet à la Cooper and Meyer (1960). There is no a priori reason why the principles by which notes group into phrases should be the same as those by which phrases make up sentences.

That extrametrical grouping intuitions may cause unlikely or dubious judgments is apparent in their example from the beginning of Mozart's G Minor Symphony (K550). Ex. 4.2 shows their grouping and metrical analysis (the hypermetric (.) is "open to interpretation", p. 22). The theory does not account for polyphony, or in general for multiple layers of events, making it impossible to consider the interference of background and foreground. The grouping therefore reflects only the events on the upper staff. Application of GPR 2 (involved with proximity) to the melodic line as a series of pitches with durations and articulation, but without meter, will produce group boundaries between notes 2-3, 3-4, 5-6, 6-7, 8-9, corresponding to the articulation slurs. GPR 1 however prescribes avoidance of very small groups, which explains why the authors do not at the basic level group the slurred notes separately from the non-slurred. This rule seems unjustified: the articulation creates a grouping into 2+1 notes (the slur implies an early lift-off of the last note under it), which coincides with the interchange of unaccented and accented sub-beats, a metrical factor excluded from grouping in GTTM.[2]

Ex. 4.2 Mozart, Symphony K550, with analysis by Lerdahl and Jackendoff (after GTTM ex. 2.10 and 3.1).

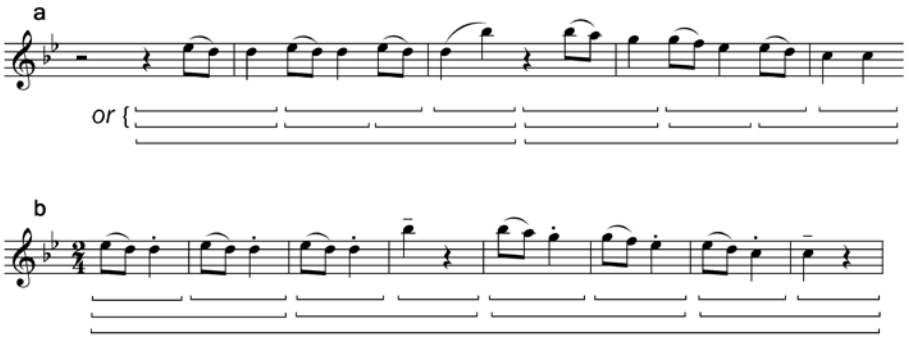


1 Anticipation is central to the theories of Meyer (1956), Narmour (1990) and Huron (2007).

2 The implausibility of GPR1 becomes apparent in the analysis of the Scherzo of Beethoven's Sonata op. 2:2 (their Ex. 2.3). On the one hand the grouping analysis takes the right hand figure and the left hand chord together; on the other hand, these two elements are discussed as two separate 'motivic cells' – which certainly makes sense, especially of we take the pianistic-gestural nature of these motifs into account.

The larger grouping may be derived from an application of GPRs 5 and 6, concerning symmetry and similarity. The status of the longer group of notes 7-10, shown between levels, is unclear; the group seems to belong to both.[1] Their neglect of the alla breve pulse, which is established in the introductory measure, produces strange results. Metrically the upbeat motif (1-2, 11-12) should include the quarter rest; it is the half note upbeat which gives the theme its propulsion. The repeat of the upbeat (notes 4-8) disrupts this slower pulse of the alla breve rhythm, and this friction may cause, against the apparent symmetry, a tendency to hear the theme as syncopated, as in Ex. 4.3a, second level: the second group is interpreted as a kind of parenthetical elaboration of the preceding upbeat. I will not insist that this hearing is uniquely correct; it seems to me however that the interference of the repeated three note motif and the alla breve meter creates an ambiguity and suspense which belies the grouping of Ex. 4.2 (it clearly betrays a hearing in $\frac{4}{4}$).[2] That the effect of meter cannot be excluded from grouping may be illustrated by transplanting the same pitches and durations into a different meter (Ex. 4.3b). Here articulation and accent are altered, because these largely depend on metrical placement. The grouping is radically symmetrical, and striking banality is the result.

Ex. 4.3 a, b



Grouping according to GTTM therefore seems to be a hybrid which includes a number of parameters while excluding the basic one, meter; it is rather like describing the folds in a robe as if there where no body underneath which sustains them. The reason is maybe the curious premise that "metrical structure consists of *beats* organized hierarchically" (p. 25) – a view that paradoxically implies that meter is *timeless*, since in their view beats have no duration. Because metrical beats are supposed to be evenly spaced, however, there can never be beats *without* associated time-spans.

4.4 Time-Span Reduction

With these objections, it becomes difficult to give a sensible interpretation to the notion of 'time-span reduction'. It is impossible to see meter on the one hand as independent from time-spans, and

1 In Ex. 4.2, as in GTTM 3.1, the group occupies an in-between level; in GTTM 7.14 the lower level. Cf. GTTM 2.3: in (a) a similarly undivided group occupies the top level, leaving a gap (non-adjacency) in the lower level; in (b) it hovers in between.

2 The time signature is omitted in all of their examples from K550.

grouping on the other hand as independent from meter. Time-span reduction would on these altered premises simply coincide with grouping reduction, implying that the analysis by assigning events to a structural hierarchy might proceed directly on the basis of grouping. This might well be the most reasonable procedure. According to GTTM however, time-span reduction involves both time-spans, for smaller units, and, confusingly, grouping, for units exceeding time-spans, even though motifs (or 'cells') as smallest groups usually inhabit time scales shorter than the bar (p. 119). It leads to a strange terminological and analytical confusion, where groups are equated to time-spans and time-spans are stretched so as to accommodate groups (p. 125, 127). With the categories of 'beginning' and 'cadence' (p. 134) the authors effectively introduce (as they acknowledge) within time-span reduction "the traditional notion of musical *phrase*" (p. 168).

In practice, this reduction turns out to be mostly harmonic. The association of such a reduction with (at the shorter time scales) metric time-spans, by the recursive selection at successive levels of one main event or 'head' (tone or chord) *per time-span,* produces curious results. In their Ex. 5.12 for instance (Mozart's A Major Sonata K331), the harmonic progression contains one harmonic scale degree in the first two measures, two (or one) in the third measure and three in the fourth. Their analysis halves the number of 'events' per measure with each reductive step: the top level contains two numberings per measure; the second level, one (selecting from measure 3 the 'non-chord' they identify as *"vi*^{7}*"*), and one for measure 4; in the third level, measures 2 and 3 drop out; and in the final reduction, only the tonic of measure 1 remains. Their *"vi*^{7}*"* is more plausibly seen as a double *Wechselnote* on V_6, and should have disappeared in the second level of reduction. Time-span reduction, by taking measures as units of harmonic analysis, obscures the harmonic rhythm and connects moments in the progression rather as if one would summarize a text by picking the first word of every sentence. It is curious that while phrasing (grouping) is separated from meter, though it depends on it, harmonic structure is here suspended from the metrical grid.

In the structural hierarchy displayed by time-span reduction, events are assigned different levels on the basis of functional dependency or dominance. Given two subsequent events *x* and *y*, one will be structurally dominant (and have a place on the next higher level), the other will be dependent or elaborative. The crucial question is: of what nature is this dependency?

In order to find a uniform quality which allows them to reduce the multiple factors determining the relations between subsequent tones or chords to one common denominator, the authors introduce the concept of 'stability'. Either *x* or *y* will be the more 'stable'. This has an apparent plausibility. We have clear preferences for a piece to end on I, not on V, on I^5_3, not I_6, on a strong beat, not a weak beat. Generalization of such considerations is less secure. Though the tonic I may be stable, I-I-I-I-I… is not, while it is hard to say at which point stability ceases. Also, it is doubtful whether we feel the same about the stability of such different phenomena as a resolution relative to a suspension, and a root position relative to an inversion. Maybe 'stability' is merely a false abstraction, pulling disparate phenomena under one umbrella. In any case, we should expect a syntax of music to elucidate exactly those nuances, as well as many other factors which make one phrase grammatical and another musical nonsense, instead of painting all music in the lighter and darker shades of grey 'stability'.

According to Lerdahl and Jackendoff, it is "the raw material of the given tonal system" that yields "a scale of stability among pitch configurations".

> Broadly, the relative stability of a pitch-event can be thought of in terms of its relative consonance or dissonance. For example, a local consonance is more stable than a local dissonance, a triad in root position is more stable than its inversions, the tonic is the most stable harmony, the relative stability of two chords is a factor of the relative closeness to the tonic (or the local tonic) of their roots on the circle of fifths, conjunct linear connections are more stable than disjunct ones, and so forth. (GTTM p. 117-8)

The concepts of consonance and dissonance are part of acoustics and psycho-acoustics; they play no part *directly* in musical syntax which is concerned with a conventional idiom. The resolution of a dominant seventh chord to a chord with its root a fifth lower is a rule of progression which may be formulated without reference to dissonance. Similarly the 'stability' of a root position relative to an inversion is in classical music a rule of the game: not a piece of natural input, but the object of chord syntax.

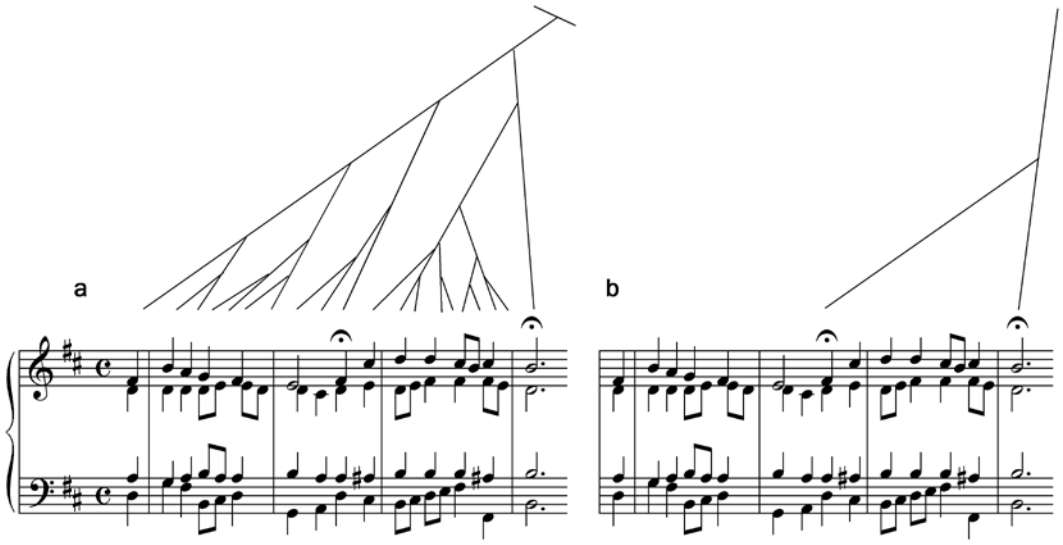
The discussions in several places of a Bach chorale setting (*Matthew Passion* Nr. 53) bring to the surface several of the problems signalled. It is the tonal harmonization of a Phrygian melody, which through various peregrinations goes back to Hassler's *Mein G'müt ist mir verwirret*. It is uncertain how this chorale setting is perceived as 'tonal music'. The fact that the same melody recurs in the Matthew Passion in various harmonizations may contribute to an awareness in the listener that what is heard is 'melody plus harmony'.[1] The consequence is that it is hard to decide what exactly constitutes an 'event' in this piece: the melody and its harmonic interpretation are only loosely joined as two distinct 'events'. Tonal hearing of this apparently simple example is not only fractured by historic awareness, but by features of the setting itself. Cadences in D major and B minor/major constantly interchange; the Phrygian finalis F-sharp, harmonized as D-major, gives no feeling of tonal closure.[2] The overarching branching in their Ex. 6.25 connects the first chord to the last, which is considered the most stable event. Though the first half phrase (measures 1-2) cadences in D, the phrase as a whole is (according to my intuition) directed to B minor, and over the chorale as a whole B minor has a stronger presence than D major. We might therefore prefer a branching different from the one proposed by Lerdahl and Jackendoff (Ex. 4.4). However, when one tries to carry through this hierarchic analysis it is hard to make out what constitutes an *event*, and which event depends upon which, on what grounds exactly. The fact that the first chord receives subsequent cadential confirmation is no reason to transfer this hierarchic position to the upbeat (an appreciation of the trajectory is part even of a 'final state' understanding: it's not all in D that ends in D). The D major chord under the first fermata seems by cadential confirmation and linear progression the first choice for a major branch (Ex. 4.4b). This branch might then connect to a

1 In GTTM it is referred to as *O Haupt voll Blut und Wunden*, but this particular setting carries the text *Befiehl du deine Wege*. Dahlhaus, *Zur Kritik der Harmonielehre* (1969), in Dahlhaus (2000 Vol. 2: 206) has criticized the didactic use of Bach's chorale settings, which in their tonal-modal mixture are ambiguous. The example is also criticized by Peel and Slawson in their review of GTTM (1984).

2 The authors are aware of tonal ambiguity (GTTM: 142-3), but fail to draw the conclusion that there is no true tonic D: "The chorale oscillates throughout between the tonic and its relative minor."

stronger branch from the B minor chord under the second fermata, which could connect (on the same level?) to the fermata chords of measures 8 (end of the repeat) and 12 (B major). It is not obvious why any of these chords should be considered more 'stable' than the others, and the chorale seems to me neither directed toward the end, nor dependent on the beginning. The dependencies which have motivated the various branches in Ex. 4.4a are not of the same kind; it is therefore doubtful that the recursive hierarchy represented by the tree structure corresponds to musical reality.

Ex. 4.4 Bach, Chorale *Befiehl du deine Wege:* (a) after GTTM, ex. 6.12, 6.25; (b) alternative



4.5 The Schenkerian Component

When an object is given a place in a hierarchy, it suffers a loss of identity. The first sounding note of a piece of music in performance looses most of its qualities as that particular sound when we perceive it as the fifth of the tonic, an upbeat in a motif, and so on. In our perception of musical phenomena, we combine certain basic shapes (gestalten, whatever they are) into larger wholes, which are different in kind. When we categorize phenomena in hierarchic fashion, what counts as an object on one level plays no part on the next higher level. There are no absolutely concrete things, and at higher levels of categorization things become ever more abstract. Psychoacoustically, pitch is a phenomenal category dependent upon vibration, within a frequency range which does not allow us to perceive it as 'rhythm'. At a certain frequency, perception 'switches' from one mode to another, from fast rhythm to low pitch. Perception in the pitch domain is however not reducible to rhythmic perception, and the rules which govern the use of pitch in music are different from those governing duration. Within the domain of tonality, pitch is reinterpreted on several levels: pitches are organized into chordal simultaneities and 'progressions', as well as melodic sequences, figures, motifs and themes. Within one level, the theories of harmony and counterpoint compete. At the next level, these harmonic-melodic events cohere within a key; the key is part of the more comprehensive system of tonality. These levels again cannot be seen simply in a part-whole relation (chords or scale

degrees 'making up' the key, keys 'making up' tonality). Tonality is not a set of 24 keys, but an abstract domain of possibilities. In between stands the composition, as a partially closed frame of reference, itself an abstraction over a wide range of phenomena (scores, performances, ideas, and so on). A composition is not a super-motif, as a river is not a super-drop, an organism not a super-cell, and any material thing is not a super-atom. Yet, a city is somewhat like a super-building, a mountain is in many respects a super-rock; and a heap of sand is exactly like the larger heap of sand into which it may disappear: there, relations are recursive. Goethe, claiming that in plant forms *Alles ist Blatt*; Riemann, who interpreted Rameau's *tout est cadence*[1] in the terms of Hegelian (or Fichtean) dialectic; Schenker, who claimed that the relations between two individual consecutive tones determine musical form at all levels; and Chomsky, who attempts to derive linguistic syntax from one basic operation *(everything is Merge)*, apply with varying degrees of sophistication the recursive simplicity of the phenomena in the second category (heaps of sand) to those which belong to the first.

In GTTM, the notion of stability serves to sustain a pervasive recursive structure. It fails to do so in a credible way because of its vagueness and arbitrariness. Schenker takes a different approach to a similar end. In his reductions concrete elements (pitches) are made to do the work of more general principles, projecting the 'foreground' phenomena of voice leading onto the basic level of organization, conceived as a far receding 'background'. We might call it 'substitution of the concrete for the abstract'.

Schenker, according to Lerdahl, "anticipated generative linguistics" by developing "a comprehensive analytic method that generates structure in a series of self-similar hierarchical levels, starting from a simple underlying form and yielding, through elaborative and transformational operations, the surface variety of a given piece".[2] The key element in Schenkerian reduction, voice leading, does however not have a very important place in time-span reduction or in GTTM as a whole. The authors reject the *Ursatz* (a rudimentary two voice schema) as generative principle, but do not deny its reality, apparently, as reductive abstraction: any tonal composition "will tend to reveal an [!] *Ursatz* at the most global reduction level; it is not necessary to posit such a structure in advance. The *Ursatz* is an effect, not a cause, of tonal principles" (p. 139-40).[3] The Schenkerian approach (and its idealist orientation) remains evident in prolongation reduction, and may also be detected in time-span reduction. It is manifest in the idea that a reduction level may be established by residues from surface events; in the representation of these residues as selected notes or pitches, instead of more abstractly, as scale degrees or functions; and in the idea that this process can be meaningfully carried on till the piece is reduced to a few notes or chords. This vision of hierarchy, though strongly focused on pitch phenomena, seems mostly inspired by the simpler hierarchy of meter: a strong beat is a beat (and nothing else) on the next higher level. By analogy, a 'stable' chord should be that same chord on the next level of reduction.

1 "Excepté le passage d'une tonique à quelque note que ce soit, tout est cadence, parfaite, rompue, interrompue ou irrégulière, en y comprenant leur imitation." Rameau (1760: 140).
2 Lerdahl, Fred. 2013. *Musical Syntax and Its Relation to Linguistic Syntax*. In Arbib 2013, 257–72.
3 Cf. Lerdahl (2009: 187).

According to Time-Span PR 9 the final cadence should generally be considered dominant relative to the beginning ("In choosing the head of a piece, prefer the structural ending to the structural beginning," p. 174). The motivation for this rule is "the intuition that tonal pieces are fundamentally goal-oriented" (ib.). This is an intuition shared with both Schenker and Riemann. It seems however strongly shaped by nineteenth century aesthetics – it has lead Riemann to his notorious rebarrings, conform the prejudice that arsis must precede thesis. There is no justification for this intuition in eighteenth century music. At the highest reductive level, it is hard to see what exactly this dependency or goal-orientation may *mean* exactly. If it means that the first movement of K331 is not really in A major unless this is confirmed by a final full cadence, we only have to transpose the cadence to perceive that we still hear a piece in A, but with a wrong ending. At this most general level we are not analyzing the structure of the piece, but the key. The key is not established by a final cadence, nor by an initial I, but by harmonic syntax.

In Schenkerian reduction, at every level a 'surface' phenomenon is retained. The motivation for this is clear: these selected pitches are connected by long range voice-leading. Within GTTM the reason is less clear. Lerdahl and Jackendoff address the issue with the following argument:

> In Schenkerian analysis one might wish to say that two identical events forming a prolongation are somehow the same event at a more background level of analysis. In the tree notation the two events would connect as a strong prolongation, but neither event would dominate the other by right or left branching. Thus there would be no progression toward tension or relaxation, but stasis. The intuitive justification for such a construction might be that at a certain level of abstraction one hears a hierarchy not of discrete events but of harmonic areas.
>
> However, this approach is inadequate. It goes against the grain of the theory as a whole, in which each event is taken in principle as 'real' at all relevant levels of analysis. One may hear harmonic areas (these are in any case enclosed within a prolongation), but one also hears all the distinctions among individual events, even those that are strongly prolonged. It is impossible to hear absolute stasis, if only because events take place in time and hence form rhythmic relationships that produce tensing or relaxing effects. (GTTM p. 184)

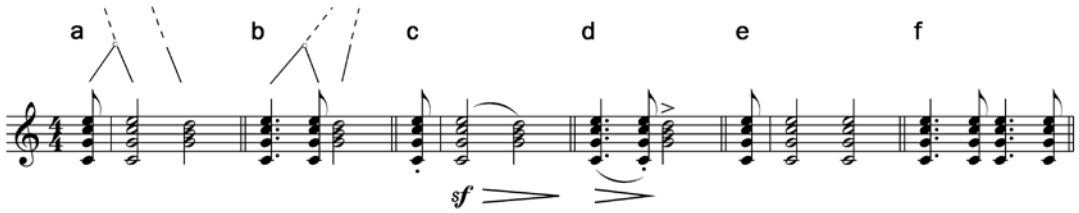
The assumption that by 'hearing' harmonic areas one would 'hear' stasis is not correct. It seems to reflect a typically Schenkerian interpretation of 'hearing', implying that if we hear, say, the main theme of K550 as within G minor (as a key area), we would hear nothing but a static G minor buzz – as if one could listen to one reduction level. We do not 'hear' the key G minor, but infer it from the manifold of 'events'. In other words, an equivocation is created by using the word 'hearing' for two different cognitive processes.

In spite of the Schenkerian outlook of time-span reduction, it does not offer an equivalent to the Schenkerian concept of *prolongation*. The authors therefore propose a complementary reduction system, which in their theoretical framework seems a kind of bonus rather than an inherent necessity. In this Prolongational Reduction, Schenker's informal 'proto-generative theory' (p. 337, n. 1) finds some degree of formalization through WFRs and PRs. The authors approach this with an air of mystery:

> In prolongational reduction [...] the domains of analysis are less predetermined, and consequently are rather difficult to explain; moreover, they cannot be explained prior to the actual building of trees. It is therefore convenient to circumvent this problem for the moment and to turn to tree building as if the problem were already solved. (GTTM p. 183)

For them, what the notion of prolongation expresses is a fluctuation of tension and relaxation, or a hierarchic composite of such fluctuations. Tension and relaxation are associated with 'prolongational anticipation' and 'prolongational repetition'. In Ex. 4.5a, the first chord anticipates the second; in 4.5b, the second is a repeat of the first. (a) Shows an instance of relaxation ("the anticipation [...] relaxes into the following down beat"), (b) of tension ("the repetition [...] tenses in response to its surrounding stronger beats") (p. 184). The reasoning, or intuition, is probably that 'anticipation' expresses or creates excitement, and that the fulfilment of the event anticipated carries an element of reassurance. However, with equal plausibility one may say that the anticipation in (a) serves to make the downbeat more exciting (tense), and that the offbeat chord in (b) effects a release. This is suggested by the dynamics and articulations as in (c) and (d). In (e) and (f) it is shown that the dotted figure of (b) might conventionally be repeated (inducing or containing little tension), while the repeat of the downbeat chord of (a) sounds emphatically static, not answering to the tension created.

Ex. 4.5 (a, b: after GTTM ex. 8.8)



This leaves doubt about intuitions of tension and relaxation associated with even very simple musical figures, stemming from the obscurity of the notion itself. For *what or who* tenses? Is this tension something inherent in musical structure, or a psychological response? If it is a response, is it voluntary or involuntary? Does it have a physiological correlate? If not, where exactly does this kind of elastic vibration take place? Mostly the authors speak of tension as something inherent in music itself,[1] but it is hard to see what it means that 'the music gets tense', other than that it moves away from the tonic or that a certain harmony fails to resolve, but than it is mere tautology.

The notion is left in rudimentary state in GTTM, but has been at the core of Lerdahl's later elaborations of *Tonal Pitch Space* (2001), a geometric projection of tonal relations (or a highly complex and sophisticated variation upon the circle of fifths). Implied in the spatial representation of tonality is the notion of distance, which is familiar to musicians and theorists in an intuitive, informal and metaphorical sense. The key of A♭ major is obviously more 'remote' from the key C major than G major, because within the system it takes more steps to get there by regular

1 "Prolongational reduction expresses one of the most basic rhythmic intuitions: the breathing in and out, the tensing and relaxing, inherent in the motion of pitch-events." (GTTM: 285). However, in the passage quoted above "tensing or relaxing effects" seems to refer to the listener.

modulation, that is, by progressions which can be interpreted within one key on a chord to chord basis. Given a mapping of all relations onto some spatial model, in two or more dimensions, distances between any two chords or keys can be calculated. The assumption is now made, that this space is not merely a theoretical construct showing relations some of which may be realized in the composition, but a kind of Newtonian absolute, pre-existent space; that the reality of this space is cognitive; and that the progress of music in time is experienced as a trajectory through this space. The varying distances are perceived by the listener with a sensation of tensing and relaxing, and this sensation may be projected onto the music as a corresponding tensing and relaxing of *the music*.

The least obvious of these assumptions is maybe that distance is perceived by measurable degrees of tension, an idea that seems to derive straight from the stimulus-response models of the earliest music psychology. It has been made the subject of experimental research by, among others, Carol Krumhansl (1996) and Lerdahl and Krumhansl (2007). The first reports on an experiment in which perceptual tension was measured by asking the subjects to push a slider. Krumhansl leaves undetermined *what* exactly is being measured (her abstract calls it a metaphor). The later report is more explicit, relying on the details of Lerdahl's newly developed pitch space. Here, they define tension as "the specific sense created by melodic and harmonic motion: a tonic is relaxed and motion to a distant pitch or chord is tense; the reversal of these motions causes relative relaxation". This is informally linked to physiological sensations:

> The expression 'tension and relaxation' also has the advantage of invoking physical motion and exertion beyond a specifically musical function. Everyone experiences physical tension and relaxation, and it is common to extend the terms to mental and emotional terrains as well. Consequently, it is relatively straightforward to ask experimental participants to respond to degrees of tension and relaxation and thereby elicit consistent interpersonal responses [...]. (Lerdahl and Krumhansl 2007, p. 329-30)

This conclusion, that it is simple to ask subjects for their responses, betrays a fatal circularity in the setup (basically the same as in the earlier experiment): one may ask the subject to respond to the music; and one may conceivably measure her response (by some physiological test, like a lie detector); but to ask the subject to respond to degrees of tension by measuring her own responses is simply begging the question. In this way, it all depends on the subject's interpretation of the assignment. The exact form in which the assignment is communicated is then crucial; and despite all its meticulous detail, the report is silent on this essential element. The fact that the subjects could interpret the assignment at all suggests that they must have had a fair idea of what the researcher wanted.[1]

Lerdahl and Krumhansl speak of tension as something inherent in music in ways awkwardly reminiscent of the animal urges in Schenker's biological *Tonwelt*.[2] Lerdahl has compared the perception of these tonal motions with the perception of interacting dots in a cartoon animation, which are anthropomorphized by the viewer ("the dots are events, which behave like interacting

1 Lerdahl (2001: 142) recognizes six different "kinds of musical tension" besides tonal tension. This raises the question how tonal tension can be measured separately.

2 "[...] music tenses and relaxes", "[...] the music tenses away from the opening I until the pre-dominant ii [...] ". Lerdahl and Krumhansl (2007: 329, 334). Cf. Schenker (1906: 6).

agents that move and swerve in time and space [...]").[1] This is misleading: musical 'events' do not behave. We do not really interpret the resolution of a leading tone as one character snuggling up with another, or sitting down with relief. This way of speaking creates a confusion that can be avoided by thinking simply of the tonal relations which allegedly give rise to this tension as musical syntax, with the kind of implication (V → I) that has tempted earlier theorists to speak of musical 'logic' (5.3 below). If we interpret a musical sentence as meaningful or expressive, it is mostly because as a whole it has the character of (somebody's) communicative action.

Graphic representations of the patterns inherent in tonality *(Tonnetze)* have been proposed before, by Riemann among others, and have become a specialized branch of music theory since the 1980's. For Riemann, it was a development of the *Musikalische Syntaxis* of 1877, and it seems indeed that harmonic (or rather, harmonic-metric) relations more than anything else in music constitute a 'syntax' remotely comparable to linguistic syntax (5.2). Lerdahl and Krumhansl point out that "The experience of tension [...] draws on prior knowledge of the abstract hierarchies [of pitch space] [...]";[2] and compare this with the kind of linguistic knowledge that makes one expect to hear *then* after hearing *if*. This expectation may, of course, also be called 'tension' and be experienced, possibly even measured as such, but it remains *knowledge* of an abstract structure that exists as a convention regardless of any speaker's sensations.[3] To put it somewhat crudely: if a listener judges, within a C major context, that the tone D is less 'stable' than G, and C♯ even less, it is *she* we may congratulate, not the music theorist. If she judges a I-V-IV-I cadence as regular, or interchangeable with I-IV-V-I, there are good reasons to say she is mistaken.

Music itself doesn't tense, and whether the listener does, doesn't matter. If aesthetic appreciation were a matter of fluctuating tensions, we might attribute an aesthetic sense to any vibrating string. At most, such sensations can be a symptom. We may either relax at the repetitive cadences of some classical symphonic finale or get tense, enjoying the noise or finding it just too much; this in itself bears no direct relation to our understanding of the syntax of dominants and tonics. Evidently, multiplying leading notes and suspensions produces an increase of tension (felt, imagined or inferred), because these are the syntactically appropriate devices for that purpose. Tonal relations are an abstract reality apart from any realization in a composition, and apart from any particular listener's knowledge and experience. As musical syntax, these relations are a vehicle for expression (by means of the totality of musical structure) rather than a stimulus, 'tensing' the listener. Modelling these relations has great value for analysis, and indirectly for the study of cognition: the model may not mirror the listener's unconscious knowledge ('competence'), but it *does* represent something she has to get a hold on, in some way. My suspicion is that this is done by bottom up processes, the application of learned schemas, rather than by orientation in an absolute pitch space (Chapter 6).

1 Lerdahl (2001: 191).

2 Lerdahl and Krumhansl (2007: 358).

3 "[...] the structure of the basic space is not merely cultural but reflects both the human impulse to hierarchize and the ear's perception of differences in consonance and dissonance [...]" Lerdahl (2001: 190).

4.6 'A Deep Parallel between Music and Language'

Though they have borrowed part of their methodology from linguistics, the authors have sensibly not attempted to create a musical grammar by reinterpreting a grammar of language. However, their avoidance of more overt parallels seems somewhat overcautious. That they were writing their book after, and in indirect response to Leonard Bernstein's Harvard lectures, *The unanswered question* (1976),[1] may explain their reticence. Bernstein's naively optimistic premise was that

> [...] there are similar functions, cognate processes operating in both music and language which are discoverable by linguistic method. All we need is to have analogous terms in which to articulate them. (Bernstein 1976: 57)

In his 1977 review, Jackendoff criticizes Bernstein's lack of intellectual rigor, and specifically his failure to keep two questions apart: that of musical universals, and that of a parallelism between the universals of language and music. "This confusion inevitably leads him to the dubious strategy of searching for musical universals by drawing analogies with linguistics".[2] In favour of Bernstein's general approach, one might argue that the separation of musical and linguistic universals, and hence musical and linguistic 'faculties', is an unwarranted assumption a priori, and that we should look rather for general musicolinguistic universals, or for linguistic universals relevant to music, in case some aspects of music are parasitic upon language. That this goal could be attained by simply applying available linguistic theory to music is, of course, mistaken. In line with the basic tenets of Generative Grammar, Jackendoff states that "The theory of a highly constrained innate linguistic ability makes a similarly constrained innate musical ability far more defensible than does a behavioristic attitude that the mind is infinitely plastic".[3] Infinite plasticity is not the only alternative to the existence of two distinct faculties or abilities, and this line of thinking is much less appealing nowadays than it may have seemed in 1977.

It is only in a kind of postscript to GTTM that Lerdahl and Jackendoff discuss similarities or parallels between music and language. One area where a bridge between the two might be created is in the use of PRs, introduced in GTTM but at that time not in use in linguistics (they make proposals to apply them in several areas of linguistics: relative scope of quantifiers, like *every, all, some*, and *many*; pragmatics, outlining communication strategies; and the semantics of categorization (the multiple meanings of 'is a …' predications). These are matters of methodology. Substantial commonalities are found however in *prosody*.

It must be noted that vocal music is not discussed in GTTM. If a Martian would study earth music through this book, she (or he, or it) would probably acquire an image of music very much in line with the 'idea of absolute music'. The only vocal music cited, the Bach chorale, is stripped of both its text and its vocality (and mistitled). With a broader, and more historically minded view of the repertoire, a reference to prosody might have been made not outside but inside the theory, as

1 GTTM: ix.
2 Jackendoff (1977: 885).
3 ib. 887.

part of the theory of phrase construction which in GTTM is transferred to the more abstract 'grouping'.

The authors limit themselves to a discussion of the stress-related phenomena of syllables and feet. Stress comes in degrees of weaker and stronger, is associated with syllables and distinct syllable types, and in many cases is distributed over small groups of two or three syllables (feet). A tree notation for stress distribution was used by Liberman and Prince.[1] Taking as an example the word 'reconciliation', we see that the six syllables of the word form three feet of two syllables, each with a succession strong-weak, with cumulative stress levels varying from *s* to *sss*. Lerdahl and Jackendoff represent these levels with the tree notation they developed for time-span reduction. Every *s* node than becomes a dominating branch in the time-span tree (p. 317). It is obvious that the same can be represented in rhythmic notation, if loosely interpreted as in free recitative (). It shows that time-span reduction is basically a metric hierarchy imposed upon harmony.[2]

One of the main attractions of Generative Grammar is, no doubt, that it has anchored a potentially free floating structural theory to a cognitive basis. It may have done too little to substantiate that basis empirically. It is unlikely, anyhow, that a hypothetical generative grammar of music could be founded on a similar basis. The universe of possible musics is not contained in our minds in the same way as (under the assumptions of Chomskian Generative Grammar) the universe of possible languages is contained in Universal Grammar. As I have argued in Chapter 1, there are several apparently independent cognitive faculties involved in music, such as pitch discrimination, sense of meter (beat entrainment), and the recognition of prosodic shapes. The pluriformity of music makes it unlikely that all three together, and possibly more, can be defined as the innate core of music cognition. Also, the structural rules of classical music are overwhelmingly conventional, as much part of the world of art as the works themselves, and subject to alteration by individuals and 'schools'.

4.7 Cognitive Approaches since GTTM

To review a work thirty years after publication would be pointless, if its reception had reached a stage were merits and deficiencies were coming clearly into perspective. Since the publication of GTTM the study of the mind has increasingly become the territory of cognitive science (as the authors remind us in their Preface to the 1996 reprint). Jackendoff has in the meantime taken considerable distance from generative orthodoxy, and particularly from its most 'hierarchically uniform' manifestation in Minimalism.[3] His *Simpler Syntax* (with Peter Culicover, 2005) levels the trees of linguistic syntax by giving up a number of traditional hard core assumptions, notably the derivation of all variety and irregularity from a underlying 'deep' structure, through transformations

1 Liberman and Prince (1977).

2 Lerdahl and Jackendoff (GTTM: 329) note that "the similarity between the theory of time-span reduction and the theory of prosodic structure seems to us to be a much more significant parallelism between music and language than has ever, to our knowledge, been pointed out." This must be facetious, since they just invented time-span reduction (cf. 4.4). The close relations between prosody and traditional phrase structure have always been recognized.

3 Culicover and Jackendoff (2005: 88-94, 540-4).

and hidden levels; and also, significantly (with an argument pointing out multiple branching in the visual domain and music),[1] compulsory binary branching. In the same book they state their faith that the alternative, bottom-up framework of Construction Grammar (see 6.1.1) "represents a deep and true discovery about the nature of language. [...] it has deep consequences for the problem of acquisition, for now the acquisition of rules can be seen as a principled extension of the acquisition of words, guided by generalizations over the contextual restrictions of individual words".[2] Yet, despite its dependence upon a by now outdated linguistic model, GTTM has not been superseded or lost its standing.

Writing about *The Capacity for Music: What Is It, and What's Special About It?* (a title evidently chosen in response to the notorious *The Faculty Of Language: What Is It, Who Has It, And How Did It Evolve?* by Hauser, Chomsky and Fitch, discussed in 3.5), Lerdahl and Jackendoff restate the basic claims of GTTM.[3] They make two basic, highly contestable assumptions: that music is a unitary phenomenon, like language; and that it has a grammar in the same sense as language has a grammar. Following the distinction of HCF between the narrow and broad definitions of the language faculty, they speak of the musical capacity broadly defined (comprising all cognitive resources applied in music), and the narrow capacity for music, which includes those resources specific to music. They have kept their faith in recursive pitch reduction.[4] They relieve the formalist tendencies of GTTM by a somewhat impressionistic overview of the emotion-related qualities of music, including narrative aspects of musical form, and "physical patterns of posture and gesture", involving, almost as an afterthought, "the deep relationship between music and dance" (without considering the possibility that sense of meter, or beat entrainment, might be involved as an essential motoric component).[5]

The most important footnote to GTTM in this article is one that has the potential for upsetting the whole theory – the recognition that music has its 'building blocks', schemas or patterns, which at various levels simultaneously may determine structure, as language has its idioms, fixed expressions, and syntactic constructions. This provides us with an alternative to the generative paradigm, which either complements it (as in Culicover and Jackendoff's *Simpler Syntax*) or replaces it (as does Construction Grammar). Since the recognition and use of such patterns makes fewer demands upon a 'narrow', primarily syntactic faculty of language (or music), it may provide us with a theoretical basis for the comparison of musical and linguistic structure (Chapter 6).

An overview of empirical studies inspired by GTTM is beyond the scope of this chapter.[6] Temperley's comprehensive *Cognition of basic musical structures* (2004) is "an attempt to quantify and

1 ib. 114.
2 ib. 536.
3 Jackendoff and Lerdahl (2006).
4 Restated in Jackendoff, *Music and Language*, in Gracyk and Kania (2011: 101-12).
5 Jackendoff and Lerdahl (2006: 65). Cf. Jackendoff's (1987: 236-238) more extensive remarks on 'body representation'. Nussbaum (2007: 47, 71), arguing that "musical experience is haptic, kinesthetic, and spatial" wants to read GTTM syntactic hierarchies as "motor hierarchies and action plans".
6 Those include Krumhansl (1990); work by Deliège (1987) is summarized in Sloboda (2005: 140). Hamanaka et al. (2006) describe a computer implementation of some of the time span rules of GTTM. They do not think of this as an exploration in music cognition, but as tool for such technologies as performance rendering and information retrieval systems.

implement Lerdahl and Jackendoff's initial conception, and to expand it to other musical domains" (p. 13). Combining formalized intuitions with an implementation through simplified musical input data (a 'piano roll transcription'), he proposes quantified preference rule systems for meter, melodic phrase structure, contrapuntal analysis, tonal-pitch-class and harmonic analysis. Though he treats his subject with considerable breadth, linguistic aspects are not strongly present.

Given the radically recursive structure of GTTM's reduction practice, it is not surprising that it has been considered from the viewpoint of linguistic Minimalism. Linguists Katz and Pesetsky have collaborated on a substantial paper with the provocative title *The Identity Thesis for Language and Music* (2011; henceforth ITLM). They rely on GTTM as "the best-developed proposal of its type, unrivalled in comprehensiveness and insight" (p. 1). With a few emendations, they argue, the theory can be aligned in such a way with linguistic theory (specifically, Principles and Parameters and the Minimalist Program) that music and language appear fully congruous in their structure (as reflected by the theory). The obvious differences between language and music are due to the fact that they deal with different materials:

> All formal differences between language and music are a consequence of differences in their fundamental building blocks (arbitrary pairings of sound and meaning in the case of language; pitch-classes and pitch-class combinations in the case of music). In all other respects, language and music are identical. (ITLM p. 3)[1]

An advantage of such a strong thesis over the cautious approach of GTTM is that it offers much better chances of bringing to light interesting exceptions, if not total falsification. On the other hand, to speak on this basis of 'identity' may seem presumptuous; rather like arguing that bread and bricks are basically the same thing, because both are baked in an oven (the only difference being one of dough and clay). Evidently, the key word is 'formal': "It is those properties of language and music that are purely *formal* that do not appear to differ" (p. 36). 'Properties of language and music' seems to mean, in the context of their theory, properties of their *grammars*, which are supposedly identical, and the step from formal identity to system identity depends on the hypothesis basic to Generative Grammar (and GTTM), that the grammar reflects the operations of the mind. Evidently music and language are not identical in the way they function; and it is not clear that this is *merely* a consequence of their building blocks, or that a formal theory without these building blocks has much substance, especially if these are arbitrarily narrowed down to "pitch-classes and pitch-class combinations". What remains is a thesis not about language and music, but about two theoretical models or grammars whose adequacy is in doubt.

As a reflection of its formal grammar, music according to ITLM is stripped of its most obvious phenomenal features. This may explain the baffling observation that "Nothing in the musical or linguistic surface [...] would suggest that the cognitive representation of tonal and rhythmic prominence in music shares the same formal mechanism as that of phrasing and stress in language"

1 Tsoulas (2010) likewise argues "that music and language share the computational system and differ in the nature of external systems and the basic units of the lexicon of each system", the computational system (syntax) operating through the minimalist principle Merge.

(p. 66). That musical phrasing is *very much* like linguistic phrasing has always been recognized, and may be explained by the long history of music's association with text. It would be surprising if musical phrasing would rely on entirely different formal (i.e. cognitive) mechanisms.

According to ITLM, prolongational reduction is the component most close to linguistic syntax (p. 3). Its binary branching shows the effects of Merge. Branches extending across a node show the head of each merged constituent, which may form relations with a nonadjacent element. In linguistic syntactic trees, the head carries a category label, which shows not only which element is the head but also transforms this into a higher level category (as when determiner and noun form a noun phrase). This, according to Katz and Pesetsky, is "merely a matter of graphics" (p. 17). They put their faith in the principle of hierarchic uniformity, or what they call the "Schenkerian conjecture", that pitch class is a constant through all levels, and thus permits "interactions at a distance" (p. 44).[1]

The authors find an omission in GTTM, a neglect to account for "the role of the cadence in establishing a key and providing closure for a passage or piece" (p. 3). Here they note an additional, close parallel with language: the full cadence is an instance of Internal Merge. Whereas regular or external Merge is the clustering of two elements, internal Merge is the attachment of one element *already merged* with a third element. It is the consequence of the stipulation that branching must be binary: any *apparent* set of three elements {a, b, c} can only be the product of recursive re-combining: {a, b} {c} → {a, _} {b, c}. In other words: in any threesome there are two pairs, one with a partner displaced. A linguistic example is so-called *wh*-movement as in (1):

(1) The audience demanded to know whose composition the pianist was playing _ .

Here the underline marks the syntactic object position *(X's composition)* implied by the verb. According to Generative Grammar, object and verb perform their obligatory syntactic Merge, but in the phonological surface the object is 'moved' from its regular place to an earlier place in order to merge with the main clause (p. 38). "Somewhat more controversially," the authors continue, "it has also been argued that Internal Merge may also move the lexical heads of larger phrases without displacing the phrase itself *(head movement)*." In their French example sentence (2):

(2) La fille n'achètera pas le livre.

Here the elements of tense and negation *(-ra pas)* intervene between verb and object. Because it is normative that verb and object are directly joined, the verb *(achète)* is both externally merged with the object *(le livre)*, *and then* internally with the verbal tense (expressed by *-ra*).

The same feature, head movement, is supposed to apply to the subdominant-dominant-tonic cadence. Since Merge operates in an strictly binary way, the string [S D T] is not allowed. The grammar therefore generates two strings [S D] [D T], in which, in the sounding surface structure, a 'movement' has taken place of D from the first string to the second: [S _] [D T]. This theoretical abstraction has no plausible relation to perception: the dominant has its place *between* subdominant and tonic, and (unlike the linguistic examples cited) it is not clear where else it could or should have

1 "The hierarchic, headed nature of GTTM's structural representations correctly predicts the central discovery of Schenkerian analysis, the existence of 'interactions at a distance'" (p. 44). In fact, GTTM trees are a *reflection* of the Schenkerian principle.

been. The cadence is not a modification of a syntactically more regular or basic form, as *La fille n'achètera pas le livre* may be supposed to have been derived from *La fille achète le livre*. It is all motivated by the obligatory binarism which both GTTM and ITLM have borrowed from traditional Generative Grammar. In music, at least, there is no strong indication that this is anything else but an externally motivated prejudice, and in linguistics it has been abandoned by many generativists (including Jackendoff).

'Merge' is a concept of high abstraction, though the word suggests something concrete. Linguistic elements are rarely 'merged' in the ordinary sense, becoming indistinguishable and inseparable. The pairing of morphemes leads to different kinds of entities from the coupling of words or phrases. 'Merge' stands for all these kinds of pair-formation, thus giving expression to the idea of hierarchic uniformity-plus-binarism. Whatever its linguistic virtues may be, it has no recognizable relevance to music.

4.8 A Generative Grammar of Music?

A theory of language may assume a common, general capacity for language, which enables humans to use it with a fair degree of complexity, if not creativity. The problem is that we have no access to the actual working of our minds, only to external data such as written and spoken discourse. Analysis of the data may allow us to find grammatical rules, but how these should be mapped onto the mind remains a major scientific challenge.

Though we may reasonably assume that there is also a generally shared musical competence ('musicality'), it is not clear how this should be defined. We measure linguistic competence by the speaker's ability to communicate in language. The kind of behaviour that may be taken as evidence of musicality is, by contrast, diverse and interpretable in many ways. If linguistic grammar is a set of rules, operations or computations which have some kind of instantiation in the mind, allowing (and urging) us to use language, it is not clear what the equivalent should be in music. Common musicality is often taken to be the ability to enjoy music, to respond to it appropriately, to recognize pieces and basic patterns – roughly, something like understanding a language without speaking it. In GTTM, the notion of musical grammar is therefore restricted to *parsing*. But as we do not measure linguistic competence by the reader's ability to understand a poem or novel, it is doubtful that we can define musical grammar by analyzing excerpts from the 'masterworks'. Taking as the actual objects of study *scores*, the authors heavily rely on both graphic representation, and on the concepts of traditional music theory, supplemented by a few principles from gestalt psychology. The gap between music theory and psychology remains unfilled, and is merely obscured by their idealist vision of the written score as reconstructed in the listener's mind, what I called the 'inner work of music'.

Lerdahl and Jackendoff admit that an inquiry into the capacity for music "ideally runs in parallel with experimental research on the real-time processing of music, the acquisition of musical competence (as listener or performer), the localization of musical functions in the brain, and the genetic basis of all of this"; but meanwhile maintain that "At the moment, the domain of formal

analysis lends itself best to exploring the full richness and complexity of musical understanding".[1] It may be doubted whether formal analysis, at present, really does demonstrate those riches in full. Nevertheless, I would agree that analysis (formal and less formal) makes an essential contribution by suggesting interesting, that is, risky hypotheses. Formal analysis is however not able to make those essential distinctions which are basic to cognitive research. In devising a grammar with cognitive, not merely music-theoretical probability, one will have to distinguish (1) what is cognitively basic, that is, innate (most likely including beat sense and some level of pitch discrimination); (2) what may be learned, by establishing some connection between basic faculties (which are presumed to be not encapsulated); (3) what is learned as convention, and therefore more open to theory and conscious judgment. Perception and judgment are hard to tell apart, and both are confused in the intuitions we may discern by introspection.

Before cognitive science opened up a broader range of research methodologies, studies in music psychology often gave the impression that musical understanding was equated with perception, music with acoustic input.[2] It seems however not unlikely that in developing musicality, much is learned in the way the lexicon of language is learned; and the borders of what is lexical in language (and therefore conventional) are not at all clear. It is not certain, therefore, that a descriptive grammar (such as a generative grammar) will be fully adequate: rules may be made by speech communities, as musical styles are created by musicians in a mostly subconscious, collaborative process of experimentation and imitation. This opens the possibility that grammaticality conditions are learned in much the same way in music as in language. It would be interesting to know how important so-called prescriptive grammars (which can never be *merely* prescriptive) are in the making of language. Musical syntax, being a syntax of artefacts, certainly involves a large portion of 'rules of play', the kind explained by textbooks of harmony and counterpoint. Such factors of 'how to do it' and 'how it is done' fall outside the domain of a generative theory.

An attempt to put a theory like GTTM to the test will have at least three options: to test (1) the intuitions which have guided formulation of the rules; (2) whether the rules reliably lead to sound musical judgment (by some intuitive standard); (3) whether listeners come to their understanding by computations conform these rules. The authors' intuitions are in outline commonplace (as they should be), in detail debatable on grounds of contrary intuitions or theoretical considerations, and sometimes they reflect theoretically informed prejudice. Comparing listeners' choices (in matters of grouping, for instance) with the outcome of the rules is therefore of little interest: this is basically comparing the authors' intuitions with those of listeners. The second option is feasible, given a weighting of preferences not included in GTTM. What matters most, I think, is (3): the question whether these rules tell us anything about the cognitive process itself, instead of its outcome. One major difficulty in such an inquiry is the absence, in GTTM, of the threefold distinction made above: between cognitively basic factors, those acquired at an elementary level by their connections, and what is conventionally learned. Besides, the rules are not devised to handle auditory input.

1 Jackendoff and Lerdahl (2006: 35).
2 Cf. Patel (2008: 259-60).

Lerdahl and Jackendoff start their presentation of GTTM with meter, but spoil the opportunity to give it its due by the incoherent notion of the durationless beat, and by their dissociation of grouping from meter. Faithful to both Schenkerism and Generative Grammar, they attempt to maximize the hierarchic aspect. The structure of classical music undoubtedly displays hierarchic features: meter, tonal organization (pitches, chords, keys), and phrase structure. Whether these are sufficiently formalizable to constitute a generative grammar remains in doubt. GTTM's separation of parameters, and the way they are rejoined at higher levels in Time Span and Prolongation Reduction, fails to do justice to what for two centuries has been rightly considered the foundation of musical structure: harmony and its metric coordination. Any attempt to upgrade traditional music theory to a theory of music should start with formalizing the theory of harmony, and analyzing its interactions with meter and grouping.

Though it steers away from *specific* language analogies, the GTTM approach paradoxically implies an analogy much *stronger* than I am inclined to uphold, an analogy inherent in the assumption that a theory of *music* (rather than a theory of *tonality*) should have the form of a (generative) *grammar*. The bypass from grammar to artwork reflects both the Schenkerian tradition (reducing the composition to the *Ursatz*) and that of Chomsky (reducing language to I-language, that is, grammar). Making this assumption, the authors avoid any clear distinction between innate cognitive factors, acquired listening behaviours and compositional principles, and in this way construe a musical 'competence' which is psychologically ill-defined.

GTTM has been developed within a formalist framework, relying on the aesthetics of 'absolute music'. As an alternative, we may adopt a functionalist perspective, which studies music and language in terms of their functions and of general principles of cognition and behaviour, rather than by hardwired, specialized algorithms. This may open the possibility of establishing a common cognitive basis, less specific and possibly beyond strict formalization, which may explain how one human creative ability may overlap with another. Within such an approach, conventional features may be incorporated as optional realizations within a broad space, defined by cognitive-functional constants. It requires no total, recursive organization relating the 'surface' to a unique generative 'background'.

References

Arbib, Michael A., ed. 2013. *Language, Music, and the Brain: A Mysterious Relationship*. Cambridge MA: MIT Press.

Bergeron, Katherine, and Philip Vilas Bohlman, eds. 1992. *Disciplining Music: Musicology and Its Canons*. Chicago [etc.]: University of Chicago Press.

Bernstein, Leonard. 1976. *The Unanswered Question: Six Talks at Harvard*. Cambridge MA: Harvard University Press.

Boden, Margaret A. 2008. "Odd Man out: Reply to Reviewers." *Artificial Intelligence* 172 (18): 1944–64.

Clarke, Eric F. 1986. "Theory, Analysis and the Psychology of Music: A Critical Evaluation of Lerdahl, F. and Jackendoff, R., A Generative Theory of Tonal Music." *Psychology of Music* 14 (1): 3–16.

Cooper, Grosvenor, and Leonard B. Meyer. 1960. *The Rhythmic Structure of Music*. Chicago: University of Chicago Press.

Culicover, Peter W., and Ray Jackendoff. 2005. *Simpler Syntax*. Oxford [etc.]: Oxford University Press.

Dahlhaus, Carl. 2000. *Gesammelte Schriften in 10 Bänden*. Edited by Hermann Danuser. Laaber: Laaber-Verlag.

Deliège, Irène. 1987. "Grouping Conditions in Listening to Music: An Approach to Lerdahl & Jackendoff's Grouping Preference Rules." *Music Perception* 4 (4): 325–59.

Gjerdingen, Robert O. 2007. *Music in the Galant Style*. New York: Oxford University Press.

Gracyk, Theodore, and Andrew Kania, eds. 2011. *The Routledge Companion to Philosophy and Music*. Hoboken: Routledge.

Hamanaka, Masatoshi, Keiji Hirata, and Satoshi Tojo. 2006. "Implementing 'A Generative Theory of Tonal Music.'" *Journal of New Music Reserch* 35 (4): 249–77.

Harvey, David. 1985. "Lerdahl, Fred and Jackendoff, Ray: A Generative Theory of Tonal Music." *Music Analysis* 4 (3): 292–303.

Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch. 2002. "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?" *Science* 298 (5598): 1569–79.

Huron, David. 2007. *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge MA: MIT Press.

Jackendoff, Ray. 1977. "The Unanswered Question by Leonard Bernstein." *Language* 53 (4): 883–94.

———. 1987. *Consciousness and the Computational Mind*. Cambridge MA: MIT Press.

———. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford [etc.]: Oxford University Press.

———. 2009. "Parallels and Nonparallels between Language and Music." *Music Perception: An Interdisciplinary Journal* 26 (3): 195–204.

———. 2011. "What Is the Human Language Faculty?: Two Views." *Language* 87 (3): 586–624.

Jackendoff, Ray, and Fred Lerdahl. 2006. "The Capacity for Music: What Is It, and What's Special about It?" *Cognition* 100 (1): 33–72.

Katz, Jonah, and David Pesetsky. 2011. "The Identity Thesis for Language and Music." http://ling.auf.net/lingBuzz/000959.

Krumhansl, Carol L. 1990. *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.

Lerdahl, Fred. 2001. *Tonal Pitch Space*. New York: Oxford University Press.

———. 2009. "Genesis and Architecture of the GTTM Project." *Music Perception* 26 (3): 187–94.

Lerdahl, Fred, and Ray Jackendoff. 1983. *A Generative Theory of Tonal Music*. Cambridge MA: MIT Press.

Lerdahl, Fred, and Carol L. Krumhansl. 2007. "Modeling Tonal Tension." *Music Perception* 24 (4): 329–66.

Liberman, Mark, and Alan Prince. 1977. "On Stress and Linguistic Rhythm." *Linguistic Inquiry* 8 (2): 249–336.

Longuet-Higgins, Christopher. 1983. "All in Theory – the Analysis of Music." *Nature* 304 (5921): 93–93.

Meyer, Leonard B. 1956. *Emotion and Meaning in Music*. Chicago: University of Chicago Press.

———. 1967. *Music, the Arts, and Ideas: Patterns and Predictions in Twentieth-Century Culture*. Chicago: University of Chicago Press.

———. 1973. *Explaining Music: Essays and Explanations*. Chicago [etc.]: University of Chicago Press.

Muns, Lodewijk. 2008. *Why I Am Not a Schenkerian*. https://www.academia.edu/7288829/Why_I_am_not_a_Schenkerian.

———. 2014. *Classical Music and the Language Analogy*. Humboldt-Universität zu Berlin.

Narmour, Eugene. 1990. *The Analysis and Cognition of Basic Melodic Structures : The Implication-Realization Model*. Chicago: University of Chicago Press.

Nussbaum, Charles O. 2007. *The Musical Representation: Meaning, Ontology, and Emotion*. Cambridge, MA [etc.]: MIT Press.

Patel, Aniruddh D. 2008. *Music, Language, and the Brain*. Oxford [etc.]: Oxford University Press.
Peel, John, and Wayne Slawson. 1984. "A Generative Theory of Tonal Music by Fred Lerdahl and Ray Jackendoff." *Journal of Music Theory* 28 (2): 271–94.
Prince, Alan, and Paul Smolensky. 1993. "Optimality Theory: Constraint Interaction in Generative Grammar." http://roa.rutgers.edu/files/537-0802/537-0802-PRINCE-0-0.PDF.
Rameau, Jean-Philippe. 1760. *Code de musique pratique*. Paris: Imprimerie Royale.
Riemann, Hugo. 1877. *Musikalische Syntaxis: Grundriß einer harmonischen Satzbildungslehre*. Leipzig: Breitkopf & Härtel.
Rohrmeier, Martin. 2011. "Towards a Generative Syntax of Tonal Harmony." *Journal of Mathematics and Music* 5 (1): 35–53.
Rosner, Burton S., and Leonard B. Meyer. 1986. "The Perceptual Roles of Melodic Process, Contour, and Form." *Music Perception* 4 (1): 1–39.
Schenker, Heinrich. 1906. *Harmonielehre*. Neue musikalischen Theorien und Phantasien. Wien; Leipzig: Universal-Edition.
Scruton, Roger. 1997. *The Aesthetics of Music*. Oxford: Clarendon Press.
Sloboda, John A. 2005. *Exploring the Musical Mind: Cognition, Emotion, Ability, Function*. Oxford [etc.]: Oxford University Press.
Temperley, David. 2004. *The Cognition of Basic Musical Structures*. Cambridge MA: MIT.
Thaut, Michael H. 2005. *Rhythm, Music, and the Brain: Scientific Foundations and Clinical Applications*. New York: Routledge.
Tsoulas, George. 2010. "Computations and Interfaces: Some Notes on the Relation between the Language and the Music Faculties." *Musicae Scientiae* 14 (1 suppl): 11–41.