Genesis and Architecture of the GTTM Project
Author(s): Fred Lerdahl
Source: *Music Perception: An Interdisciplinary Journal*, Vol. 26, No. 3, Music and Language (Feb., 2009), pp. 187-194
Published by: University of California Press
Stable URL: http://www.jstor.org/stable/40286102
Accessed: 01/02/2011 14:01

Genesis and Architecture of the GTTM Project

Fred Lerdahl
Columbia University

I examine the intellectual and music-theoretic origins of *A Generative Theory of Tonal Music* (Lerdahl & Jackendoff, 1983) and review the crucial steps in theory construction that led to its overall architecture. This leads to a discussion of how shortcomings in GTTM motivated developments in *Tonal Pitch Space* (Lerdahl, 2001). I conclude with a diagram that encompasses the major components of the expanded GTTM/TPS theory.

Received May 27, 2008, accepted September 22, 2008.

Key words: genesis, generative, grouping, tonal tension, pitch space

The 25th anniversary of the publication of *A Generative Theory of Tonal Music* (Lerdahl & Jackendoff, 1983; abbreviated GTTM) induces me to reflect on the ideas and decisions that shaped our project and set the course for subsequent developments.

Noam Chomsky changed the intellectual landscape of the 1960s with his reformulation of linguistic theory as the formal study of the human capacity for language (Chomsky, 1965). In the early 1970s, Ray Jackendoff and I concluded that music might be studied in similar fashion. Our interest was not in a literal transfer of linguistic to musical concepts, as Leonard Bernstein (1976) attempted. Rather, it was Chomsky's way of framing issues that attracted us: the supposition of specialized mental capacities, the belief that they could be studied rigorously by investigating the structure of their outputs, the distinction between an idealized capacity and its external and often accidental manifestations, the idea of a limited set of principles or rules that could generate a potentially infinite set of outputs, and the possibility that some of these principles might be unvarying beneath a capacity's many different cultural manifestations.

To make this vision concrete, we needed to focus on a particular musical idiom, yet with a view to how particular formulations might be generalized. We chose Classical tonal music because it was a well-theorized idiom that we knew well. In keeping with the American music-theoretic climate at that time, we began our collaboration with a consideration of Heinrich Schenker's analytic system, which in its final and most influential incarnation could be viewed as a proto-generative theory (Schenker, 1935/1979). Schenker posited an originating structure for all tonal music, the pre-rhythmic *Ursatz* (fundamental structure), comprising a I-V-I harmonic elaboration of the tonic triad supporting a stepwise melodic descent to the tonic. From this structure he developed, through stages of hierarchical elaboration and transformation, successive levels of musical structure until a musical surface, or piece of music, was reached. Crucial to his theory was that the same elaborative and transformational principles apply recursively at all levels. These structural levels and transformations could be interpreted from a late 20th-century perspective as representing off-line cognitive structures and operations. The form of his theory seemed comparable to Chomsky's generative method from S-node ("sentence node," or top level of the phrase-structure hierarchy) through phrase-structural and transformational rules to surface structure. Figure 1 summarizes the parallel; rectangles signify structures, ellipses rules. Schenker's writings, however, were rather informal. It seemed a logical first step to formalize his methodology.[1]

For a number of reasons we soon gave up this Chomskian-Schenkerian approach. First, we could not justify the *Ursatz*. Although this *a priori* construct was understandably central to Schenker, a thinker steeped in 19th-century German philosophical idealism, its status made little sense to a modern, scientifically inclined American. Nor could schema theory in cognitive psychology (Neisser, 1967) defend this construct, for the *Ursatz* is too remote from a musical surface to be picked up and organized by a listener who is not already

[1]Milton Babbitt's student Michael Kassler (1963) had already tried to cast Schenkerian theory as a propositional analytic system, and James Snell (1979) was soon to pursue a variant of the quasi-Chomskian approach sketched here. Also see Sundberg & Lindblom (1976) for a non-Schenkerian, quasi-Chomskian treatment.

Music Perception volume 26, issue 3, pp. 187–194, issn 0730-7829, electronic issn 1533-8312 DOI:10.1525/mp.2009.26.3.187

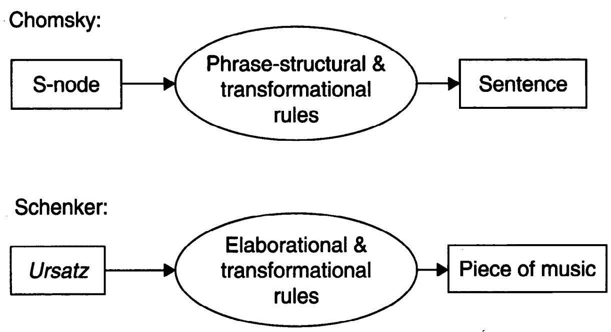

FIGURE 1. Analogy between Chomskian and Schenkerian theories.

predisposed to find it. Second, the *Ursatz* is inapplicable to music of other times and cultures. We sought a theoretical framework that could accommodate diverse musical idioms. Third, the non-rhythmic character of the *Ursatz* presented a formal and musical problem. How was rhythm to be introduced into the derivation, and why should it have inferior status? Fourth, even supposing that the *Ursatz* or some comparable structure could be justified as a foundation, there would be very many possible ways of generating a given musical surface. What would determine an acceptable derivation? How could derivations be constrained? Fifth, it was not interesting to begin our work by considering abstract background musical structures and presumed transformations; the exercise felt too abstract. Sixth and most importantly, what of psychological interest would there be even if we managed to build a system that generated this or that piece from an *Ursatz*-like foundation? What mattered to us was not the output per se but the structure attributed to the output. It was not clear how generating a piece could reveal much about mental structures and their principles of organization.

Reflecting on these considerations, we decided to formulate our theory in the opposite way. Rather than begin with a putative ideal structure and generate musical surfaces, we would begin with musical surfaces and generate their structural descriptions, as shown in Figure 2. There would be no hypothetical point of origin oriented toward a particular musical style, no systemic elevation of pitch over rhythm, and no problem of exploding derivations. We could deal directly with music of interest and value. The rules would be motivated psychologically and would represent cognitive principles of organization. The structural descriptions would correspond to predicted heard structures. The theory would in principle be testable.

Three methodological perspectives borrowed from generative linguistics helped launch the enterprise.

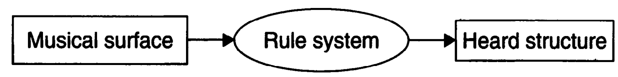

FIGURE 2. Overall form of the GTTM theory.

First, we would assume as given the musical surface—essentially quantized pitches and rhythms with dynamic and timbral attributes—without worrying about the complex perceptual mechanisms that construct the surface from the audio signal. Second, our quest for cognitive principles would proceed from our own musical intuitions. Only later would we seek experimental corroboration. Third, we would build a final-state rather than processing theory, on the view that it was advantageous to specify the mental structures in question before trying to articulate how they operated in real time. These positions were not meant to denigrate psychoacoustics, experiment, or processing. Rather, they were strategic decisions in theory construction. They also had the advantage of keeping our project within the bounds of the music-theory tradition, since most music theories take pitches and rhythms for granted, appeal to intuitive plausibility, and do not restrict the study of musical works to their real-time unfolding.

We began by developing rules to assign hierarchical structure to pitch events. The idea, sketched in Figure 3a, was to articulate stability conditions that decide the relative hierarchical importance of events—the more stable the event, the higher its position in the hierarchy. From Schenkerian theory we loosely adapted the term prolongational reduction to signify hierarchical pitch connections. (In a music-theoretic context, the term "reduction" means hierarchy; subordinate events are "reduced out.") We modified tree structures from linguistic theory to represent prolongational structures. There was no attempt to seek analogs to parts of speech or syntactic phrases. A branching pattern denoted elaboration in an expanded sense of ordinary melodic/harmonic embellishment. Thus in Figure 3b, event e2 is an elaboration of event e1; in Figure 3c, event e1 is an elaboration of event e2.

This initial attempt foundered on two related grounds. First, the theory did not specify which events

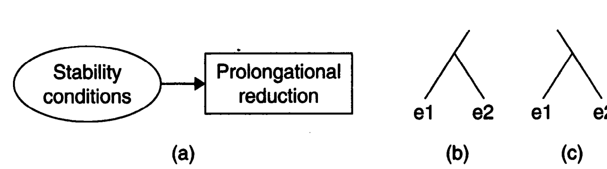

FIGURE 3. Beginnings of model of hierarchical pitch-event structures: (a) the initial framework; (b-c) the tree notation signifying elaboration.

to compare for connections. Events could hardly branch indiscriminately. The theory required nested regions of analysis. Second, even if we assumed a context as simple as a phrase or period, events often repeat within that unit. In a standard period, for example, the antecedent and consequent usually begin with an identical I, and they state the same V at or near the end of each phrase. How could the theory distinguish repetitions and assign each occurrence its proper value? On both grounds, the importance of an event could not be separated from its position in the temporal order. We could not build a rule system to assign a hierarchy of events without first developing a theory of rhythm.

The prevailing attitude in the 1970s was that pitch was well understood but rhythm was intractable. Nevertheless, some theorists had been theorizing about rhythm, either in terms of metrical groupings (Cone, 1968; Cooper & Meyer, 1960) or metrically oriented pitch reduction (Komar, 1971; Lewin, 1974). We reviewed these approaches and forged solutions to problems that beset them. The key was to disentangle grouping and meter from one another and to treat them as independent but interactive components. Once grouping and metrical structures were assigned, the two components could recombine to form a time-span segmentation that would locate each event within the hierarchical time structure. Each event would have a unique temporal address. Our rhythmic components took the form of Figure 4. (For simplicity, here and in the following flow charts I do not show the component of the musical surface that serves as input to the structural components.)

Investigation of the grouping component brought us to Gestalt psychology, for which there was an extensive literature on visual grouping. It was unclear how to make a grammar out of the Gestalt principles of proximity and similarity. Working through many grouping and metrical analyses, we found that the phenomena were gradient rather than categorical. If different factors converged on the same result, a grouping or metrical intuition was strong. If they conflicted, the intuition was weak or ambiguous. Standard linguistic grammars, in contrast, yielded yes-or-no grammatical results; there were no shades of gray. Soon we concluded that

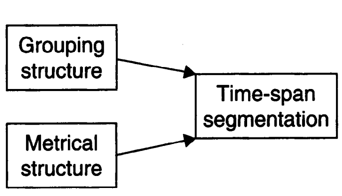

FIGURE 4. Rhythmic components of the theory.

perceptual judgments were gradient not only for grouping and meter but also for most aspects of musical structure. We decided to overhaul the entire rule system in progress. Well-formedness rules defined possible hierarchical features and replaced the function of recursive phrase-structure rules, which as a rule type we had at first imported from linguistic theory. Transformational rules, which were then central to generative linguistics, were demoted to handle marginal cases such as grouping overlap and metrical deletion. We put front and center a new kind of system, preference rules, whose purpose was to select from the many possible well-formed descriptions those few that best matched, in gradient fashion, intuitions in response to a given musical surface.

When GTTM was published, preference rules were criticized for not behaving in proper grammatical fashion (e.g., Peel & Slawson, 1984). First of all, they were seen as insufficiently quantified. While true, this was not an objection in principle. GTTM deliberately avoids the intricacies of rule quantification in order to concentrate on the musical and cognitive principles at hand. Our formal goal was not full quantification, which could come later (as it has begun to, for example, in Temperley, 2001, 2007), but enough precision for the proposed components and rules to be investigated empirically.

A deeper objection was that preference rules did not assign structure through a cycle of derivational steps. In retrospect, however, the convergent (instead of derivational) character of preference rules turned out to be a pioneering aspect of what is now a common computational framework in cognitive science, whether in the form of schema theory, parameter variables with dynamic programming, neural nets, machine-learning techniques, or Bayesian probabilities. In linguistics, they show up in phonology in optimality theory (McCarthy, 2001), in semantics in word meaning and categorization (Jackendoff, 1983), and in syntax, at least implicitly, in parametric settings of features across languages (Baker, 2001). In general, our version of preference rules occupies a conceptual midpoint between neural nets, in which a given structure arises from the strongest activation in a network without reliance on rules per se, and optimality theory, in which rule derivations are ordered and ranked according to a winner-take-all principle. From our perspective, neural nets are unable to produce enough structure, and rule rankings rigidify the assignment of structure beyond what can be justified empirically.

The development of the grouping and metrical components not only yielded preference rules but also showed the way to a novel theoretical architecture.

Rather than construct our theory around a central component from which other components are derived, as was the case for Schenker and Chomsky, our theory took the form of several independent components that coordinate with one another to generate an overall structural description. Neither pitch nor rhythm has priority; rather, they interact. This perspective has affected Jackendoff's subsequent work in linguistics, in which he posits that phonology, syntax, and semantics are independent systems whose structures are coordinated by interface rules (Culicover & Jackendoff, 2005; Jackendoff 2003).

With a rhythmic theory, a rule system, and a conception of the overall architecture in hand, we turned again to pitch structure. At issue was how to incorporate rhythmic information into a decision procedure for assigning pitch-event hierarchies. Ideas of rhythmic reduction, which could be traced back to Schenker (1921-23), were in the air. We took particular notice of Lewin's (1974) version, in which events within a given metrical frame were selected, proceeding strictly from the bottom up.[2] His method, however, lacked a grouping component or any special treatment of cadences, causing awkward results. These shortcomings could be rectified, first by replacing metrical frames with the time-span segmentation, which combines grouping and metrical features; and second by the device of cadential labeling, which preserves the structural role of cadences up to the largest level for which they function. Thus we arrived at the component of time-span reduction. Each span has a dominating event, or head, together with one or two subordinate events. The head of one span competes for dominance against another head in an equivalent span at the next larger level; and so on up to global levels.[3] The winner at any given time-span level is determined primarily by the stability conditions mentioned previously.

We found that analyses generated by the time-span component were useful but insufficient. The methodology often yielded unconvincing event sequences at underlying levels, and the regions of analysis were excessively tied to the rhythmic segmentation. Most critically, the analysis did not assign specific connections between events, especially nonadjacent ones such as between a tonic and its return in a reprise. We could informally draw trees and graphs that better captured the sense of prolongational reduction with which our enterprise began, but we still did not know how to derive or interpret such analyses.

Three steps enabled us to overcome this impasse. First, we reconceived prolongational branching in terms of nested patterns of tonal tension and relaxation. A right branch now represented an increase in tension, a left branch a decrease in tension. This interpretation happily removed our thinking from the Schenkerian shadow and gave prolongational analysis a distinctly psychological flavor. We established three node types for prolongational tension connections: strong prolongation for exact repetition (no increase or decrease in tension), weak prolongation for modified repetition (a small increase or decrease in tension), and progression for non-repetition (a relatively great increase or decrease in tension). The second step was to give up the preconception that there ought to be just one kind of pitch reduction. We posited two kinds, time-span and prolongational, as complementary ways to understand hierarchical event structures. Time-span reduction represents the hierarchy of events in relation to the rhythmic structure, and prolongational reduction represents hierarchical patterns of tonal tension and relaxation. The third step was to derive prolongational structure from global to local levels of the associated time-span reduction. We found that the best results could be obtained through a slightly relaxed correspondence between the two reductions. More precisely, time-span level ${}_{ts}L_n$ provides the events available for attachment at prolongational level ${}_{p}L_n$. If, however, and event can be found at ${}_{ts}L_{n-1}$ that can attach as a strong prolongation at ${}_{p}L_n$, that event is raised and so attached; but an event at ${}_{ts}L_{n-2}$ is inaccessible to ${}_{p}L_n$. We call this relationship between the two reductions the interaction principle. Figure 5 illustrates it schematically. It claims that a determining factor in prolongational analysis is an event's stability in the rhythmic structure. Put differently, the interaction principle acts as

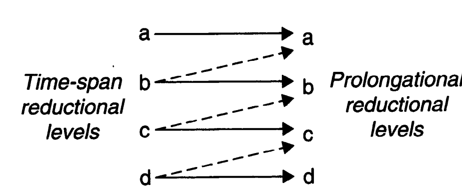

FIGURE 5. Schematic illustration of the interaction principle. A solid arrow represents a direct mapping from events in a given level of time-span reduction to an equivalent level in the associated prolongational reduction. A dashed arrow represents the access of events at a given time-span level for connection at an immediately larger prolongational level if (and only if) they can connect as strong prolongations.

[2]Although Lewin's (1974) paper was never published, its methodology can be deduced from the analyses in Lewin (2005), chapters 5 and 6.

[3]This way of thinking arose in part from Jackendoff's (1977) concurrent work on X-bar syntax, in which a syntactic phrase has a head that dominates other constituents in the phrase.

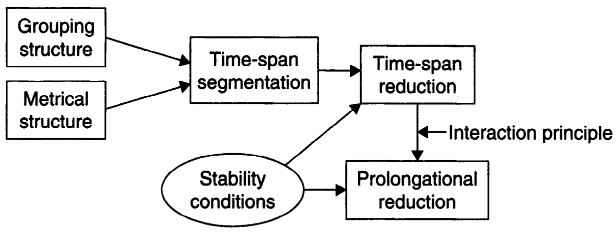

FIGURE 6. Flowchart of the components of the GTTM theory. (Subsidiary feedback arrows are not shown.)

a filter: if an event is in a rhythmically quite subordinate position, it cannot attach at global prolongational levels. Since GTTM's publication, I have found the interaction principle to apply without modification not only to diatonic tonal music but also to highly chromatic tonal and atonal music (see my *Tonal Pitch Space*; hereafter TPS) and to poetry.[4]

A prolongational analysis requires as input not only an associated time-span reduction but also the same stability criteria that feed into the time-span reduction. Combining this factor with Figures 3a and 4, we arrive in Figure 6 at a general picture of GTTM's components. This flowchart shows, in broad outline, how GTTM unifies rhythmic and pitch components within one framework. The keys to this unification are time-span reduction and the interaction principle.

Looking back on how Jackendoff and I developed the GTTM theory, I would single out three stages as crucial. First is the distinction between grouping and meter. Second is the notion of preference-rule systems. Third is the interaction principle. Of these innovations, music theorists have mostly embraced the first[5] but ignored the other two. Preference rules and the interaction principle both concern the derivation of perceived structure by rule, an area in which the music-theoretic community has shown little interest, even though in other respects music theories can be quite formal. Preference rules involve issues of cognition that lie outside the scope of most music theory. The interaction principle is problematic for some theorists precisely because it rationalizes the derivation of prolongational reduction. Music theorists commonly see prolongational analysis as belonging to the Schenkerian tradition, with its assumption that analysis is primarily an artistic activity carried out by conscious choice of the analyst. Our view, in contrast, was that if prolongational reduction is to be a meaningful construct, it must represent an aspect of listeners' unconscious understanding of music. Hence it must be systematically related to the heard musical surface. GTTM's rule system—and in particular the interaction principle—makes this systematic relation explicit. The interpretation of prolongational structure as the unconscious cognition of nested patterns of tension and relaxation gives this structure a psychological and implicitly emotive interpretation.

Music psychologists did not take long to test various grouping and metrical rules (e.g., Deliège, 1987; Palmer & Krumhansl, 1990), and they have welcomed preference rules as a tool for the study of musical behavior. The interaction principle itself may still be too technical to be tested empirically at this point; but its larger context, the perception of hierarchical pitch structures, is a topic of considerable interest to music psychology. GTTM's prolongational component, however, presents difficulties in this regard. Quantification of predictions of the metrical and time-span components can be accomplished by counting metrical or time-span levels (Palmer & Krumhansl, 1987), and grouping preference rules are amenable to quantification (see Frankland & Cohen, 2004, for a partial effort in this direction; also Temperley, 2001). But the prolongational component as set forth in GTTM is not quantifiable. Its branching types describe degrees of tension and relaxation only qualitatively. And GTTM gives merely a verbal sketch of the stability conditions that underlie both the time-span and prolongational components.

Soon after the publication of GTTM, I made several attempts to quantify the prolongational component with a view to making it amenable to empirical research. The solution began to take shape with the development of pitch-space theory (Lerdahl, 1988; TPS), which quantifies the most important of the stability conditions through computational modeling of empirical data on the tonal hierarchy (Krumhansl, 1983, 1990). The idea is that the cognitive distance of an event from a given reference point measures the instability of that event in relation to the reference point. On the assumption that the listener unconsciously seeks the most stable construal of a musical passage, TPS's principle of the shortest path selects events yielding the smallest available distances from superordinate events at each stage of prolongational reduction.[6]

[4]For a chromatic example, see the Scriabin analysis in TPS, chapter 7. My approach to the sounds of poetry treated as music is sketched in Lerdahl (2003), but only the fuller account in Lerdahl (2008) includes applications of the interaction principle.

[5]See, for example, Kramer (1988) and Rothstein (1989).

[6]Lerdahl & Krumhansl (2007) gives a summary account of how pitch-space distances, prolongational stability, and tonal tension are derived and quantified.

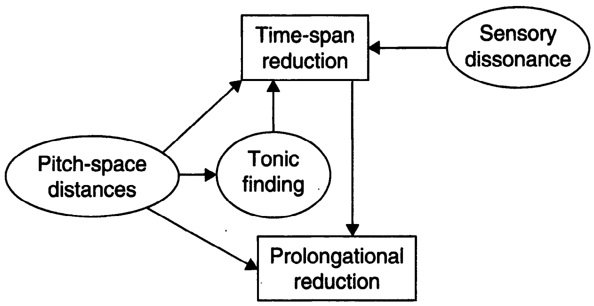

FIGURE 7. TPS's delineation of GTTM's stability conditions, together with the reduction components.

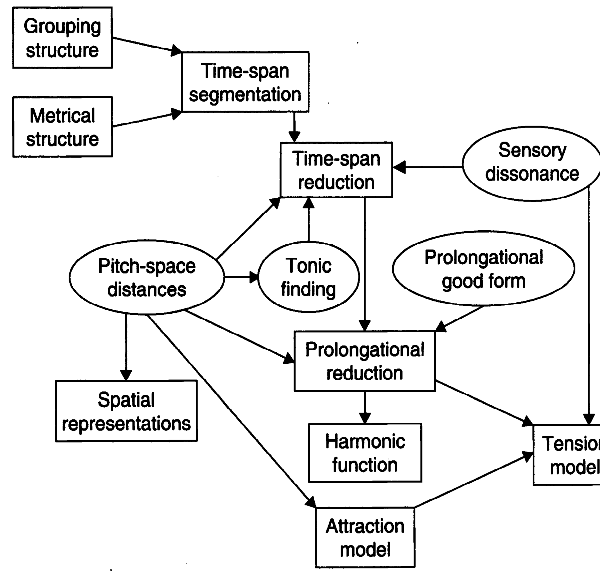

FIGURE 8. Flowchart of the components of the GTTM/TPS theory.

The principle of the shortest path presupposes the establishment of global and local tonics as the reference points from which other events are directly or indirectly measured. Therefore the theory requires a tonic-finding algorithm. Further, GTTM and the pitch-space model assume traditional distinctions between harmonic and non-harmonic tones, with the latter heard as more unstable than the former, and between chords in root position and in inversion, again with the latter as more unstable than the former. These distinctions ultimately rest on psychoacoustic grounds, so the model must include a subcomponent of sensory dissonance. Figure 7 incorporates these factors into the flowchart, for the moment leaving the rhythmic components out of the picture. The input identified as stability conditions in Figure 6 is here replaced by three inputs: pitch-space distances, tonic finding, and sensory dissonance.

Instability can also be interpreted as tension, and thus we return to GTTM's conception of prolongational structure as representing nested patterns of tension and relaxation. The tension of an event is measured in large part by its pitch-space distance down the branches of the prolongational tree from the global tonic.

TPS's more inclusive view of tonal tension demands the addition of three more factors. First is sensory dissonance. Second, to assure a smooth flow of tension and relaxation, prolongational branchings are regulated by principles of prolongational good form. These principles encourage recursive branching patterns, balance between right and left branching, and a rise-and-fall tension pattern at the phrase level. Third is the phenomenon of tonal attraction. Unstable pitches metaphorically seek to move to, or are attracted to, proximate stable pitches (Bharucha, 1984, 1996; Larson 2004). This factor extends to voice leading and harmony; it is also a partial measure of expectancy. For example, the leading tone is strongly attracted to, and expected to arrive on, the tonic pitch, whereas there is little tendency or expectation for it to move to the dominant pitch. The tension of attraction is complementary to the tension of pitch-space distance and must be part of a full model of tonal tension.

In this connection, TPS employs a kind of rule not found in GTTM: numerical algorithms. These rules are used, not yet with uniform success, for pitch-space calculations (very accurate), sensory dissonance (a good approximation), tonic finding (fair but incomplete success), and attraction (in need of revision).[7] TPS's frequent turn to this type of rule reflects a change not in attitude but in subject matter. Where the topic is pitch schemas, algorithmic rules are applicable, reflecting the constancy of long-term knowledge of the tonal system. Pitch space, tonic location in the space, and degrees of attraction within the space all concern schematic representations and operations. Where the topic is event hierarchies, however, preference rules remain requisite, reflecting the gradient interaction of diverse features.

Taking all these factors together, we arrive in Figure 8 at a flowchart of the extended GTTM/TPS theory. At the upper left are the rhythmic components of grouping,

[7]Lerdahl & Krumhansl (2007) justifies the accuracy of pitch-space distances and the approximation of sensory dissonance, and it discusses weaknesses in the attraction subcomponent. Lerdahl & Seward (2008) refines the tonic-finding method proposed in TPS, chapter 5, and notes remaining weaknesses in the tonic-finding computations.

meter, and time-span segmentation. The nexus of the chart is time-span reduction, which takes as input time-span segmentation, sensory dissonance, pitch-space distances, and tonic finding. Via the interaction principle, time-span reduction then contributes, along with pitch-space distances and prolongational good form, to the construction of prolongational reduction. To give a more complete picture, the chart adds two TPS components that have not been under discussion: spatial representations, which are mapped from pitch-space calculations; and harmonic function, which is read off the prolongational tree. Finally, the prolongational component represents not the final stage of tension and relaxation but serves as input, along with the components of sensory dissonance and attraction, to the expanded tension model.

Empirical research generally supports the GTTM/TPS theory. To take a few instances, the preference-rule behavior of local grouping rules is well established (Deliège, 1987; Frankland & Cohen, 2004). Metrical studies routinely assume GTTM's distinction between grouping and meter and the grid-like character of meter (Large & Palmer, 2002). The independence of rhythmic and pitch components is confirmed both cognitively (Palmer & Krumhansl, 1987) and neuropsychologically (Peretz & Coltheart, 2003). There has been extensive research using priming to explore prolongational hierarchies and expectancy (e.g., Bigand & Pineau, 1997). Of particular importance are studies that validate the tension component (Bigand, Parncutt, & Lerdahl, 1996; Krumhansl, 1996; Lerdahl & Krumhansl, 2007; Smith & Cuddy, 2003). As displayed in Figure 8, this component relies directly or indirectly on almost all the other components in the theory. To the extent that the tension model holds up, then, so provisionally does much of the overall theory.

My interest in continuing the GTTM/TPS project lies in a number of directions, of which I mention four here. First, I would like to improve the attraction component and clarify its role in musical expectation. Second, I would like to tighten the entire theory to the point that it is implemented computationally, that is, so that from given inputs it automatically predicts preferred outputs. Third, I have attempted to extend the theory into the domains of atonal music (TPS, chapter 8), timbral organization (Lerdahl, 1987), and the sounds of poetry (Lerdahl, 2003, 2008). I would like to flesh out these attempts. Fourth, I hope to understand more deeply the relationship between musical emotion and the ebb and flow of tonal tension.

Author Note

Thanks to Ray Jackendoff and two anonymous reviewers for suggestions that improved this paper.

Correspondence concerning this article should be addressed to Fred Lerdahl, Department of Music, 621 Dodge, Mail Code 1813, Columbia University, New York, NY 10027. E-MAIL: awl1@columbia.edu

References

BAKER, M. C. (2001). *The atoms of language.* New York: Basic Books.

BERNSTEIN, L. (1976). *The unanswered question.* Cambridge, MA: Harvard University Press.

BHARUCHA, J. J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology, 16,* 485-518.

BHARUCHA, J. J. (1996). Melodic anchoring. *Music Perception, 13,* 383-400.

BIGAND, E., & PINEAU, M. (1997). Global context effects on musical expectancy. *Perception and Psychophysics, 59,* 1098-1107.

BIGAND, E., PARNCUTT, R., & LERDAHL, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception and Psychophysics, 58,* 5-141.

CHOMSKY, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

CONE, E. T. (1968). *Musical form and musical performance.* New York: Norton.

COOPER, G., & MEYER, L. B. (1960). *The rhythmic structure of music.* Chicago: University of Chicago Press.

CULICOVER, P., & JACKENDOFF, R. (2005). *Simpler syntax.* New York: Oxford University Press.

DELIEGE, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl and Jackendoff's grouping preference rules. *Music Perception, 4,* 325-360.

FRANKLAND, B. W., & COHEN, A. J. (2004). Parsing of melody: Quantification and testing of the local grouping rules of Lerdahl and Jackendoff's *A generative theory of tonal music. Music Perception, 21,* 499-543.

JACKENDOFF, R. (1977). *X-syntax: A study of phrase structure.* Cambridge, MA: MIT Press.

JACKENDOFF, R. (1983). *Semantics and cognition.* Cambridge, MA: MIT Press.

JACKENDOFF, R. (2003). *Foundations of language.* New York: Oxford University Press.

KASSLER, M. (1963). A sketch of the use of formalized languages for the assertion of music. *Perspectives of New Music, 1*, 83-94.

KOMAR, A. J. (1971). *Theory of suspensions.* Princeton: Princeton University Press.

KRAMER, J. (1988). *The time of music.* New York: Schirmer.

KRUMHANSL, C. L. (1983). Perceptual structures for tonal music. *Music Perception, 1*, 28-62.

KRUMHANSL, C. L. (1990). *Cognitive foundations of musical pitch.* New York: Oxford University Press.

KRUMHANSL, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception, 13*, 401-432.

LARGE, E. W., & PALMER, C. (2002). Perceiving temporal regularity in music. *Cognitive Science, 26*, 1-37.

LARSON, S. (2004). Musical forces and melodic expectations: Comparing computer models and experimental results. *Music Perception, 21*, 457-498.

LERDAHL, F. (1987). Timbral hierarchies. *Contemporary Music Review, 1*, 135-160.

LERDAHL, F. (1988). Tonal pitch space. *Music Perception, 5*, 315-350.

LERDAHL, F. (2001). *Tonal pitch space.* New York: Oxford University Press.

LERDAHL, F. (2003). The sounds of poetry viewed as music. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 413-429). New York: Oxford University Press.

LERDAHL, F. (2008). *Poetry as music.* Unpublished manuscript.

LERDAHL, F., & JACKENDOFF, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

LERDAHL, F., & KRUMHANSL, C. L. (2007). Modeling tonal tension. *Music Perception, 24*, 329-366.

LERDAHL, F., & SEWARD, R. (2008). *Toward a computer implementation of the GTTM/TPS analytic system.* Unpublished manuscript.

LEWIN, D. (1974). *Analysis of Schubert's song "Morgengruss."* Unpublished manuscript.

LEWIN, D. (2005). *Studies in music with text.* New York: Oxford University Press.

MCCARTHY, J. J. (2001). *A thematic guide to optimality theory.* Cambridge: Cambridge University Press.

NEISSER, U. (1967). *Cognitive psychology.* Englewood Cliffs, NJ: Prentice-Hall.

PALMER, C., & KRUMHANSL, C. L. (1987). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 116-126.

PALMER, C., & KRUMHANSL, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 728-741.

PEEL, J., & SLAWSON, W. (1984). Review of *A generative theory of tonal music. Journal of Music Theory, 28*, 271-294.

PERETZ, I., & COLTHEART, M. (2003). Modularity of music processing. *Nature Neuroscience, 6*, 688-691.

ROTHSTEIN, W. (1989). *Phrase rhythm in tonal music.* New York: Schirmer.

SCHENKER, H. (1921-1923). *Der tonwille* (Vol. 1.). (W. Drabkin, Ed.). New York: Oxford University Press.

SCHENKER, H. (1979). *Free composition* (E. Oster, Trans.). New York: Longman. (Original work published 1935)

SMITH, N. A., & CUDDY, L. L. (2003). Perceptions of musical dimensions in Beethoven's *Waldstein* Sonata: An application of tonal pitch space theory. *Musicae Scientiae, 7*, 7-34.

SNELL, J. L. (1979). Design for a Formal System for Deriving Tonal Music. Master's thesis, SUNY at Binghampton.

SUNDBERG, J., & LINDBLOM, B. (1976). Generative theories in language and music description. *Cognition, 4*, 99-122.

TEMPERLEY, D. (2001). *The cognition of basic musical structures.* Cambridge, MA: MIT Press.

TEMPERLEY, D. (2007). *Music and probability.* Cambridge, MA: MIT Press.