An Artificial Intelligence Approach to Tonal Music Theory

James R. Meehan

Technical Report #124A

An Artificial Intelligence Approach
to Tonal Music Theory

James R. Meehan

ABSTRACT. There are virtually no computer models of tonal music theory, largely due to a preoccupation with syntax that comes from traditional linguistics, particularly from transformational grammar. Narmour's recent book, Beyond Schenkerism [5], defines Schenker analysis as a transformational system, refutes the central parts of that theory, and suggests an alternative theory of analysis. This paper draws a parallel between some of Narmour's ideas and current work in natural language processing, Schank and Abelson's Knowledge Structures [11]. The principal correspondence is between Narmour's "style forms" and the AI notion of semantic primitives. It may now be possible for music theorists to share the philosophy and methodology of AI researchers in producing programs to compose and analyze tonal music.

Technical Report #124A

August 1978, revised June 1979

Department of Information and Computer Science
University of California, Irvine
Irvine, California 92717

An Artificial Intelligence approach to tonal music theory

James R. Meehan

Introduction. Although computers are commonly used in the composition and performance of contemporary music, there is very little computer research on models of tonal music theory, for either analysis or composition of music. This seems surprising since tonal theory is in the standard music curriculum and shares many aspects of natural language, a favorite target from the earliest days of computer science, and since computer programs in Artificial Intelligence (AI) are now commonly used as process models, particularly in areas rich in symbol manipulation. There are programs that use little or no standard theory to compose nursery tunes [7], rounds [9], and even cowboy songs [8]. The early literature shows a strong influence of information theory and theories of composition based on Markov chains, weighted probabilities of state transitions, generate-and-test models, and so on. Few authors of such systems made claims of generality or extensibility, and indeed, no such system has caught on. There are very few analysis programs, good, bad, or otherwise, the principal exception being a 1968 paper by Terry Winograd [15].

If there are such strong parallels between music and natural language, what's missing?

In a review of books on computers and music [14], Barry Vercoe of MIT said:

> We seem to be without a sufficiently well-defined "theory" of music that could provide that logically consistent set of relationships between the elements which is necessary in order to program, and thus specify, a meaningful substitute for our own cognitive processes.

On that same point, Andy Moorer [4] wrote in 1972:

> ... any attempts to simulate the compositional abilities of humans will probably not succeed until in fact the musical models and plans that humans use are described and modeled.

Schenker. The currently dominant theory of tonal music is that of Heinrich Schenker (1868-1935) who defined a transformational system for music analysis long before Noam Chomsky did the same for linguistics [1]. Very briefly, the transformations reduce groups of notes on one level to single notes on the next higher level, in a fashion not unlike common parsing techniques for context-free grammars. The higher-level notes are said to be "prolonged" by the lower-level notes; a C at one level, for instance, might be prolonged by stepwise motion (C-D-E-F-G). These reduction rules theoretically apply at all levels. Finally, one is left with a two-voice structure known as the Ursatz, for which there are three possibilities: the melodic part (Urlinie) may descend a third (scale degrees 3-1), a fifth (5-1), or an octave (8-1). The Ursatz itself is a prolongation of the triad. Melodic (horizontal) motion is thus viewed as a temporal expansion of harmonic (vertical) structure. Of the many structural levels, three are distinguished: the _foreground_, which is the surface representation; the _middleground_; and the _background_, which is the _Ursatz_.

The Schenker theory has greatly enhanced our understanding of musical structure by relating the harmonic and melodic aspects of music. But as widely as the theory is accepted and taught, it remains incomplete, imprecise, and a constant subject of debate by music theorists. Two major works have appeared recently that discuss the Schenker theory from a "modern" point of view. The first is a paper by Fred Lerdahl and Ray Jackendoff [3], who have attempted to formalize music theory, improving on Schenker, from within the paradigm of generative Transformational Grammar. The second is a book by Eugene Narmour, _Beyond Schenkerism_ [5], in which the author discusses numerous weak points in the Schenker theory in particular, and in Transformational Grammar in general. He also proposes a new way to look at music. Neither of the two new approaches is yet complete; both sets of authors promise upcoming books in which all the details will be worked out. Yet the two "revisions" of Schenker are utterly antithetical, a tribute, if nothing else, to Schenker's influence.

The linguistic approach. In Transformational Grammar, one seeks a correspondence between deep structure and surface structure of sentences, and the issue of grammaticality is paramount. Similarly, in Schenker analysis, one seeks a correspondence between the background and the foreground, already knowing what transformations are possible. If we're interested in the phenomenon of understanding, then from the fact that people often make perfect sense of sentences that am ungrammatical, we can conclude that grammar isn't very important. Likewise, not even musicians are troubled by having failed to detect the <u>Ursatz</u> when they hear a Beethoven symphony.

Parallels drawn between music and language often center around the issue of syntax, and linguistics seems to be a natural choice for a discipline in which both music and language might be studied. Computer science has formalized the notion of syntax, and the problem of syntax-driven translation of programming languages is now a technological skill taught to undergraduates. The problems of natural (human) language, however, are not solvable by such syntactic methods, and the split between the "computational linguists" and the traditional linguists has widened. In computer science, the area called Natural Language Processing shares very little with the philosophy and methodology of Linguistics.

<u>Natural Language Processing</u> (NLP). Narmour's theory, called the implication-realization model, is of special interest to Artificial Intelligence researchers because it bears a strong resemblance to recent models of natural language processing, particularly the work of Schank,

Riesbeck, and Abelson [10,11]. In their view, the key to processing language is to look beyond syntax and to represent meaning via semantic primitives, relying on the notion that, at the literal level at least, you can get most people to agree on the meaning of simple sentences such as "John gave Mary a book" and "John gave Mary a kiss"; while the syntax of those two sentences is identical, the meanings are not even close. Roger Schank developed a set of "primitive acts" to describe everyday, physical actions. The theory, Conceptual Dependency, has been used in various computer programs that understand newspaper stories, make inferences, translate, paraphrase, summarize, answer questions, and write stories. Neither the text-analysis programs nor the text-generation programs rely on grammars of English syntax.

Higher-level semantic structures. Years of work with computer systems modeling these semantic primitives led Schank and Abelson to organize structures above the level of individual actions, particularly those that deal with aspects of human problem-solving. Routine sequences of actions, such as what one normally does in a restaurant, are described as scripts. Above scripts (in the problem-solving hierarchy) are plans, which involve more choices and decisions. In theory, some scripts "evolve" from plans by learning from repeated experience; while most people understand the principle of shelving books in a library, for example, and could figure out what to do if confronted with such a task, their initial behavior would likely differ from that of skilled library workers. Plans are driven by goals, which can be permanent (e.g., staying healthy), temporary (mowing the lawn), cyclic (eating food), and so on. Finally, themes account for behavior associated with roles (fireman), interpersonal relationships (spouse), and lifestyle (jet set). Understanding language, in this view, requires an understanding of people; it has much to do with cognitive psychology and very little to do with grammar.

Expectation-based parsing. The programs that parse English into these Knowledge Structures are based on expectations of meaning. For example, suppose we've seen "Mary took John ..." as the beginning of a sentence. The Conceptual Dependency representation of that, so far, would be (PTRANS (ACTOR MARY) (OBJECT JOHN)) where PTRANS is the primitve act for physical motion and ACTOR and OBJECT are conceptual case names. We can now make several kinds of predictions, each expressed in the form "if X then Y." Since we've just been parsing a noun phrase ("John"), there's a pending expectation for the word "and": if the next word is "and," then predict another noun phrase and extend the OBJECT-filler. Another prediction is that if we see a noun phrase that has the property of being a physical object ("Mary took John a book"), then make the current OBJECT-filler the TO-filler and make the referent of the noun phrase the OBJECT-filler: (PTRANS (ACTOR MARY) (TO JOHN) (OBJECT BOOK)).

The predictions for prepositions and adverbs are very diverse, and many of them depend on the higher-level semantic structures (contexts) described by the scripts, plans, and goals. For example, if the context is wrestling, then we can make a special prediction for the word "down" (a specific wrestling maneuver), entirely different from the more neutral ("default") prediction about accompaniment ("Mary took John down to the harbor"). Likewise, one can establish predictions to handle the following phrases: to task, to Los Angeles, to the cleaners, to mean that ..., into her house, into the firm, up on his offer, up the hill, on, etc.

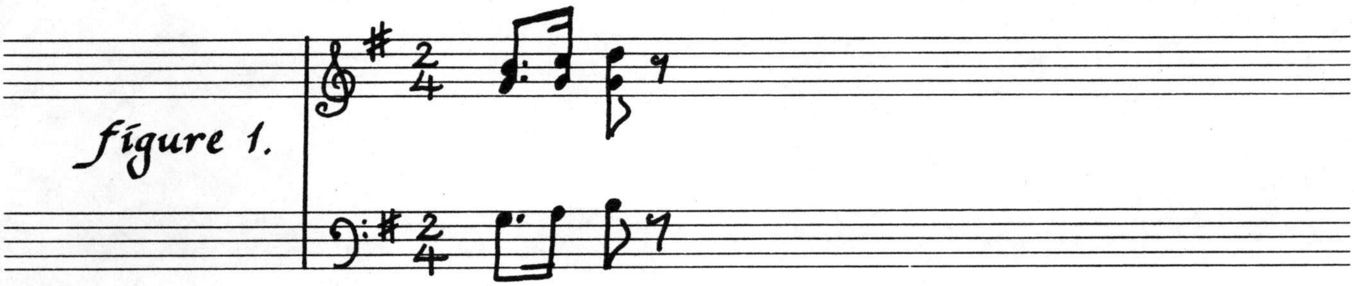
While these contexts are useful in parsing different sentences, they also make predictions about the different interpretations of identical sentences. Compare the effects of the reply "John has a cold" to the two questions "Where's John?" and "Which of you tenors has suddenly turned tone-deaf?" Music theory texts often make the same point about pitches, chords, and sequences of chords: interpretation depends on context.

Although these expectations deal primarily with the parameter of meaning, as opposed to focus or, for that matter, rhythm (e.g., dactyls in "Mary took John to the home of the president"), there is no reason in principle why the technique could not be applied to other parameters as well. AI has yet to tackle such other parameters in language in as well-organized a fashion, defining higher-level structures and coordinating the expectations from different parameters.

Narmour's theory of music. Is it possible, then, to define an expectation-based procedure for analyzing music? Narmour seems to think so. (There is a difference in terminology between Narmour and the AI researchers. To Narmour, the term "expectation" means "prediction with absolute certainty," which is not the AI sense. As I understand it, his implications are possible consequences, and realizations are actual consequences.)

In an example discussed at length in his book, Narmour describes the following implications after the first three chords of Schumann's "Soldier's March" (figure 1).

figure 1.



The next soprano note may be an E (B-C-D-E). If so, the next bass note may be a C (parallel 10ths). The soprano may eventually reach a rhythmically accented high G (B-D-G triad). The next chord may be more stable than the current I6 (I -> IV), and so on. The musical "semantic primitives" are what Narmour calls style forms,

> those parametric entities in the piece which achieve enough closure [local explanation] to enable us to understand their intrinsic functional coherence

> without reference to the functionally specific, intraopus context from which they come. [5,p.164]

They are patterns that "make sense" by themselves, and they are associated with the parameters of music: melody, harmony, rhythm, meter, etc. When we say that a certain piece is more interesting harmonically than rhythmically, for example, what we mean is that the rhythmic patterns are simple, whereas the harmonic patterns are not. Bach four-part chorales, for example, exhibit less variety in rhythm than in harmony.

Recursive rules and high-level representation. A major point of disagreement that I have with both the Narmour system and the Lerdahl-Jackendoff system is something on which they surprisingly agree. They each describe analysis as a process that uses rules recursively, implying that music analysis starts from notes and applies rules to transform these into other notes, which are further transformed by the same rules, on and on. The idea has a certain mathematical appeal to it, but is unlike the situation in NLP, where sentences are not "reduced" to other sentences, but are represented in terms of an interlingua, corresponding to no particular human language. Inferences, for example, are not keyed by specific English sentences but rather by the representation of meaning. Otherwise, all synonymous sentences would have to be listed explicitly, which is not simply inefficient but also psychologically dubious.

Only if the language for high-level musical structures IS the same as for the "foreground" will it be possible to use the same rules ("recursively" in Narmour's usage) for further analysis, although there is no particular evidence, much less a guarantee, that this should be so. It is precisely this characteristic that Schenker uses in order to reach the Ursatz. It predisposes you to reduce the number of structures you transform into, although, as Narmour has corrected me [6], it does not require that the analysis terminate in any specific top-level set such as Schenker's three Ursaetze.

To me, the lower levels of Schenker analysis are the most convincing, the explanations of the local connections between melody and harmony. By the same token, some of the Lerdahl-Jackendoff theory seems usable, particularly their preference rules. It is much more useful to know the principles of metrical grouping, for example, than to see a transformation that enforces conformity with the data structure of trees, which seems an exercise in notational convenience, at best.

Differences. Since even the simplest music has aspects of melody and harmony, it is more complex than the natural language that can currently be processed by machine. There are good AI models of what one might describe as environments rich in problem-solving behavior. That is, NLP efforts have concentrated first (and necessarily so) on meaning, and they work well on ordinary prose such as newspaper stories. What will they do with literature, either prose or poetry? They cannot yet recognize literature because there's no representation of those domains that define literature as something beyond simple prose. That's not a criticism of the Schank-Abelson work; they never set out to process art. For that matter, nor has anyone else in AI. Of course, just as there are programs that compose "nursery tunes," there are programs that produce "poetry," too, but they're on equally weak foundations. It's hard enough to model the "literal meaning" domains, and logically, that must precede the modeling of other parameters or knowledge domains. The richness of the real world, even in simple texts, makes that problem very hard indeed.

Simple tonal music, in contrast, would seem to have more domains but less complexity within each domain. Musical "scripts" abound (e.g., cadences). The few, common meters provide "solutions" that make rhythmic expectations very simple. (Of course, just as in language, one aspect of what it means to be interesting is to avoid the easy solutions, which may explain why so much of the popular music of the 1950's, with endless repetitions of I-vi-IV-V7, is mind-numbing.) Integration of domains is a more obvious problem in music than in language, but I believe that the similarities exist. Narmour, among others, points out the differences between music and <u>the linguists' view</u> of language, and I certainly agree with him. It is not at such a low level as syntax that the two are similar, and I find Smoliar's comparison of the language of computer programs and the language of music [12] inappropriate.

Another difference in theories of music and language is the synchronic/diachronic contrast. In any system, it is tempting to seek a set of factors that, with one set of values, describes one historical point or style, with another set of values, a different point, and so on. You can then invent a theory about the nature of the changes from one set to the next, and you claim you have a diachronic model, an epistemological philosopher's stone. Narmour hopes to do the same with compositional styles: what distinguishes Beethoven's Fifth from the Sixth? In natural language processing, this is viewed as part of the problem of learning. It's hard enough, right now, to model how people <u>use</u> knowledge structures, for instance; modeling how they <u>acquire</u> them is a higher-level problem, requiring a synthesis of all the experience gained from building many individual models. In other words, it's too soon to expect any significant answers about The Big Problem, the "universals" of music. One seriously doubts Lerdahl and Jackendoff's claim:

> Preliminary investigation has indicated that the theory can be modified to produce structural descriptions of pieces in styles as diverse as Macedonian folk music, North Indian music, and 14-century French music, by changing various specifics of rhythmic and pitch structure. [3,p.166]

Composition. A language parser is not also a language generator, even though the two may share a theory of meaning. Current generators start with a representation of the meaning of what is to be said, as provided by a paraphrase program, an inference mechanism, an event simulator, or a question-answerer, for example. That is, they don't start from scratch. In fact, they make the need for higher-level structures (e.g., problem solvers) even more apparent.

There are musical composition "tasks" that provide some direction, such as the harmonization of a given melody. As any music student knows, musical knowledge is as easily tested by composition as by analysis, and the same holds true, certainly, in NLP.

Conclusion. Where does an AI person start, then? With a music theory textbook, perhaps, but reading it with the task in mind of representing the information there in a computer program. To what end? A theory of music should certainly explain aspects of the undeniably tonal music with which we are daily bombarded, such as the Bee Gees or Barry Manilow. Even the 5-note theme from Close Encounters of the Third Kind is (alarmingly) tonal. But my own preference is to concentrate on harmony and melody -- it is difficult to model even the simplest tonal music without them -- and, to a lesser extent, rhythm, with an initial goal of writing chorales. This is the type of experiment in choosing semantic primitives, forms of representation, and control structure for which there is ample precedent in AI. What music theory lacks is not the concept of expectations or semantic primitives, but rather the organization and detailed specification of such concepts, which would lead to higher-level information structures and reasonable process models for analysis and composition. If our analogy with research in natural language processing is valid, such a knowledge-based system will provide better results than previous attempts.

ACKNOWLEDGEMENTS

This paper is a revision of ICS Technical Report 124. I would like to thank Eugene Narmour, Otto Laske, Stephen Smoliar, the ACM reviewers, and the many correspondents who made valuable suggestions for improving it.

REFERENCES

1] Noam Chomsky.
Syntactic Structures.
Mouton, The Hague, 1957.

2] Otto E. Laske.
Understanding the behavior of users of interactive computer music systems.
Revised version. School of Music, University of Illinois, Urbana, Illinois, 1977.

3] Fred Lerdahl and Ray Jackendoff.
Toward a formal theory of music.
Journal of Music Theory 21(1):111-172, Spring 1977.

4] James Anderson Moorer.
Music and computer composition.
Communications of the ACM 15(2):104, February 1972.

5] Eugene Narmour.
Beyond Schenkerism.
University of Chicago Press, 1977.

6] Eugene Narmour.
Private Communication, 31 October 1978.

7] R. C. Pinkerton.
Information theory and melody.
Scientific American 194(2):77, 1956.

8] H. Quastler.
Discussion, following Mathematical theory of word formation, by W. Fucks.
In E. C. Cherry, editor, Information Theory -- Third London Symposium, Academic Press, New York, 1955, p. 168. Cited in Joel E. Cohen, Information theory and music, Behavioral Science 7, 1962, p. 163.

9] Gary M. Rader.
A method for composing simple traditional music by computer.
Communications of the ACM 17(11):631-642, November 1974.

10] Christopher K. Riesbeck and Roger C. Schank.
Comprehension by computer: Expectation-based anaysis of sentences in context.
Research Report 78, Yale Computer Science Department, October 1976.

11] Roger C. Schank and Robert P. Abelson.
Scripts, Plans, Goals, and Understanding.
Lawrence Erlbaum, Hillsdale, New Jersey, 1977.

12] Stephen W. Smoliar.
A Parallel Processing Model of Musical Structures.
PhD dissertation, MIT, 1971.

13] Barry Truax.
A communicational approach to computer sound programs.
Journal of Music Theory 20(2):227-300, Fall 1976.

14] Barry Vercoe.
Review of Harry B. Lincoln, The Computer and Music, and Barry S. Brook, Musicology and the Computer.
Perspectives of New Music 9(1)-10(2).

15] Terry Winograd.
Linguistics and the computer analysis of tonal harmony.
Journal of Music Theory 12(1):2-49, Spring 1968.