Duilio D'Alfonso[*1]
[*]*Conservatory of Music, Latina, Italy*
[1]duilio.dalfonso@tin.it

Dynamic Grammar of Tonality: The Real-Time Understanding of Tonal Music

ABSTRACT

In the field of tonal cognition studies, the *Generative Theory of Tonal Music* (Lerdahl and Jackendoff 1983) is the milestone from which a plenty of theoretical and empirical research flourished in the last decades, aimed at deepening aspects of the theory. A main issue regarding GTTM, particularly its prolongational theory deriving time-span and prolongational reductions, is widely recognized: the approach is static, a 'final state' approach, and the structural descriptions are given for the entire musical passage under analysis, without considering how such structures are inferred during the real–time process of music understanding. Indeed, the so-called preference rules, for deriving prolongational reductions, are hardly conceivable as a real-time parsing system. Thus I will introduce a 'Dynamic Grammar of Tonal Music' (DGTM), developed as a real-time parsing system inspired by the 'dynamic turn' in linguistics. DGTM is devoted to modelling tonal music understanding, especially the interaction between stored context and expectancy generation and changing, during the listening process. In order to deepen its cognitive implications, the model will be introduced by some considerations on the music/language similarities and differences, in the general perspective of formal linguistics. The grammar of tonal syntax will be presented as a categorial grammar for generating parsing trees for chord progressions, in a left-to-right, step-by-step fashion. The inferential device of the grammar is employed to formalize expectation as forward-looking inference. In the conclusion, a more general psychological interpretation of the model is proposed.

1. INTRODUCTION

Notwithstanding the magnitude of studies and theories on the grammar of tonal music from the eighteen century on, only in the last five decades the issue has been approached from the cognitive side. Traditional teaching methodologies of harmony, counterpoint, and composition (consider, for example, the Neapolitan school of 'Partimento', see Sanguinetti 2012), as well as the Schenkerian theory of tonality, are basically normative, aimed at codifying and transmitting a historical compositional practice. Only with the cognitive turn in linguistics and psychology in the 1960s of the last century, a turn then imported in theoretical musicology, this situation began to change.

The listener naturally acquainted with the tonal idiom, by means of mere environmental exposure, has become the goal of investigation, and the focus of theoretical and experimental research has been moved on the mental processes underlying the comprehension of tonal music. We hereafter use the expression 'tonal cognition' to intend the sense of a tonal context induced by a sequence of tones melodically and/or harmonically organized according to the grammar of the tonal idiom. In other words, tonal cognition is focused on the mental representation of the key, as a representational (subconscious) level in the real-time musical experience.

The psychological reality of such representation has been proved and tested by several experiments: for example, the various probe tone trials witnessing how different tones fit differently in an established (by a chord, a scale, a cadence) tonal context (Krumhansl and Kessler 1982), or the experiments showing the so-called priming effects of chords, an evident 'facilitation' in the cognitive processing of chords when related to a tonal context (Bharucha 1987). Moreover, tonal cognition has been investigated not only behaviorally, but also by means of neurophysiological technique (see Maess *et al*. 2001; Fitch *et al.* 2014 for a review on the topic).

1.1 The Main Perspectives on Tonal Cognition

Once ascertained the psychological, and neurological, foundation of tonal cognition, several formal models of the cognitive capability underpinning its mental processing have been developed. Probably, the best known is the key-finding algorithm of Krumhansl and Schmuckler (described in Krumhansl 1990). A Bayesian version of this algorithm has been then advanced by Temperley (2007). The crucial aspect of these procedures is the *statistical* approach: the frequency distribution of pitch-classes in a tonal piece is compared with the so-called key-profiles of all the 24 keys (the fitting rate associated with each tone in a given tonal context, roughly corresponding to the average distribution of occurrences of tones in a musical phrase in a key), and the key better matching the input distribution is selected.

Directly inspired by the 'connectionism' in cognitive science, Bharucha (1987) proposed a neural-computational, and sub-symbolic, model for tonal cognition, arguing that there is no symbolic layer at the interface between sound perception and the emergence of meaning in music cognition (see also Bharucha *et al*. 2012). In its neural network, as pitches are perceived, according to a given distribution, the activation spreads from tones to chords and from chords to keys, coherently with a schematic representation listeners have interiorized by means of the persistent exposure to the tonal music environment. More recently, Chew introduced a geometrical model, derived from Longuet-Higgins' works on music cognition in the (traditional) perspective of artificial Intelligence (Longuet-Higgins 1987): the so-called 'spiral array' (see Chew 2014 for an updated overview of this theory), in which tones are spirally arranged around a cylinder, triads are represented by surface points equidistant from the triadic components, and key activations are mapped into central points in the cylinder, equidistant from the triads points.

Ultimately, these models are not strongly alternative: they share the basic idea that tonal cognition is due to some sort of tones distribution. But the problem of distributional or activation-based models (either statistical or geometrical) is: they undermine the relevance of the order of pc-events. These systems are affected by the ‘indifference’ to the order. As a consequence, they completely neglect expectancy, while, as I will argue, the dynamics of expectations is an essential feature of tonal cognition, and of its mental processing.

Moreover, ordering of events in tonal music, like in natural language, is due to the linearization of hierarchical structures. Thus, a formal model of tonal cognition should account for the linearization of hierarchical representations, as well as for the recovering of hierarchical structures from a linear sequence of events.

1.2 The Role of Expectation

Expectancy is an essential feature of musical experience, involving different parameters, as widely witnessed by the musicological literature (from Schenker 1979 to Salzer 1962). There is also a significant literature in cognitive musicology and psychology of music (from Meyer 1956 to Narmour 1990 and Huron 2006), devoted to demonstrate the role of expectation in music perception, as well as its neurophysiological and evolutionary correlations.

Expectancy is obviously a crucial aspect of tonal cognition, since it is active not only at an ‘infrastructural’ (somewhat unconscious) level of music perception, but it is at work at the subconscious level of the emotional meaning of tonal music. Specifically, a well-formed harmonic progression, in experienced listeners, generates the prediction of its proper continuation. Such predictions can be satisfied or violated. I will return on the important issue of harmonic expectancy’s violation for the emotional responses to musical stimuli (see Margulis 2006 for an investigation of these implications in melodic expectation). In the meantime, we can identify the basic psychological function of a chord progression with its *expectancy generation/change potential*, i.e. with its capacity of modifying the ‘mental predictive state’ of the listener. Indeed, tonal sense is more ‘implicative’ than ‘adaptive’: for example, given a chord progression, it consists on the representation of the ‘preferred’ continuations, if any, of that progression.

Now, ordering and hierarchical structure of harmonic events are what generate expectation in tonal music. Hence, a model for tonal cognition aimed at formalizing its mental processing should be able to represent a hierarchy of ordered units.

1.3 Prolongational vs. Context-Free Grammatical Model for Tonal Harmony

Prolongational trees in GTTM account for hierarchies and long-distance dependency between pc-events, i.e. of prolongational relations. A similar job is accomplished by a context-free grammar for tonal harmony, such as the model of Rohrmeier and Neuwirth (2015).

Compare the two structural descriptions of a chord progression in Figure 1. On the left, there is the prolongational tree analysis, on the right the parsing tree in the context-free grammar style. As one can see, context-free analysis implies a categorization of harmonic units — the terminals vocabulary of the grammar is a string of scale degrees, while prolongational tree is not applied to labelled units but directly to pc-events. This difference is somewhat relevant, since prolongational trees do not presuppose a previous application of some ‘chords labelling’ procedure. Indeed, the tree on the left in Figure 1 can be interpreted as a description minimizing distance/tension (see Lerdahl 2004). So interpreted, prolongational trees do not presuppose any sort of categorization of the chords progressions: they directly capture the listener feeling of tension/relaxation patterns, the tension growing produced by distances and the recovery of stability, when distances are reduced.

Since listeners do not recognize single chords as such (as presupposed by context-free grammars), at a pure psychological level, prolongational description is to be preferred to phrase structure description. Nevertheless, it is always possible to provide a prolongational interpretation of the parsing tree released by a context-free grammar. Along this way, we can preserve the cognitive relevance of the prolongational tree, within a well-understood mathematical formalism. The use of context-free grammars allows both flexibility and the computational control assured by the production rules. Indeed, it is well known how such formalism is the theoretical basement for parsing and generation systems design (in computer science).

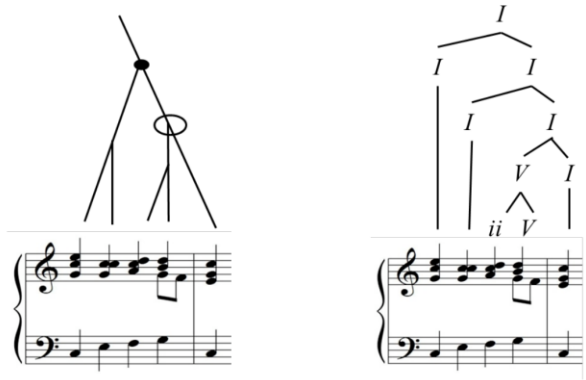

Fig. 1. Prolongational vs. context-free representation of the structure of the expanded cadence.

In the categorial, and dynamic, model of tonal syntax I will propose, a kind of trade-off between prolongational formalism and context-free model will be pursued, just to avoid giving up the properties (psychological the former, formal and computational the latter) of both the theoretical frameworks.

1.4 Language vs. Music Syntax

Before going through the presentation of such model, let me briefly, and more generally, consider the crucial issue of the language/music syntax similarity, correlated with the hierarchical organization of musical syntax, sometimes viewed as problematic. We can start with the following question: to what extent is tonal syntax really context-free? (But, marginally, one can wonder to what extent natural language syntax is context-free.) And then, to what extent is tonal syntax hierarchically featured? Now, hierarchical representations, in linguistic theories, account for:

1. Recursive expansion of categorial units (recursive External Merge, in the minimalist jargon);
2. Movements (or Internal Merge, for dealing with long-distance dependency).

These phenomena cannot be dealt with linear, or sequential, models (for example, with a Finite State Automaton): they require a hierarchical representation of the syntactic structure. The first one because of the unpredictability of the nesting depth. The second one because movements are constrained in a way depending on constituent structure — regarding phrasal units or the heads of a phrase. Now, it is quite unquestionable the musical syntax exhibits recursive expansion of functional units, while more problematic is the issue of constituent movements (the point will be sketched later). My point here is that musical syntax and linguistic syntax share the essential property of recursive, and not predictable in depth, expansion of categorial units. Consider Table 1.

T′			
I	T		
I	D		I
I	S	d	I
I	IV	V	I
I	IV ii	V	I
I	IV V/ii ii	V	I
I	V/IV IV V/ii ii	V	I

Table. 1. Predominant area expansion.

The predominant area is progressively expanded, and, at least theoretically, the magnitude of this expansion is not predetermined. One can mention the so-called *indugio*, a schema of the Gjerdingen's Schemata Theory (2007), characterized by a wide prolongation of the ii^{56} chord (a typical predominant chord in the classical style, often used by Mozart).

Then, where does the hardness of conceiving musical syntax as hierarchical organized rise? Why is it so often disputed? A possible answer: hierarchical representations, with unpredictable expansion of constituents, require the necessity of introducing a categorization of musical units. Now, such categorization cannot be based on semantic aspects of lexical units, as it happens in natural language. Instead, it might be anchored to functional aspects of chords, as they occur in a complete cadential progression. While in natural language formal and functional aspects are interlocked in the lexicon, in tonal music the functional level is determined only in chords concatenation, i.e. at the syntactic level. There isn't in music a pure formal level for which a chord receives a tonal function, exclusively determined by its inherent structural properties. It is well-known, for instance, that the I degree chord, in root position, plays the role of the functional (initial or closing) tonic when occurring in the appropriate position, but it can also be a passing chord when occurring in other contexts. The function of chords, in tonal harmony, is a strict contextual property, while in language the syntactic function of words is already somewhat constrained in the lexicon, where words are assigned a grammatical category. This is a strong, or maybe the strongest, difference between music and language, and, probably, the main reason why it is so difficult to formalize musical syntax, especially in a hierarchical fashion. Anyway, the functional categorization of constituents and its hierarchical organization, in musical syntax, are strictly related: they imply each other.

Ultimately, although the issue of the (perceivable) hierarchical construction in tonal syntax is disputable, the alternative, a flat model of tonal syntax, presumptively conceived as a 'Markov chain' — where each chord affects at most the following one, is anyway problematic. One can argue for different ways of listening to tonal music: while a way of listening involves only the feeling of the chord passing into the following one, maybe appraising the so-called 'parsimony' in the voice-leading (Cohn 1997), a more trained listener can have some intuition of the hierarchical structures of a tonal progression, with the consequent grasp of the function of chords, their hierarchical and recursive prolongation, and of long-range tensing/relaxing patterns.

2. THE TONAL TEMPLATE

Following Katz and Pesetsky (2011), constituents' movement showed by language syntactic constructions (interrogatives, relative clauses, etc.) also involves musical syntax. Particularly, they advance a theory of 'Head Movement' of the dominant chord in the full cadence: it is attached to the following tonic, and dependent to it, although it is in turn the head prolonged by prefixed predominant harmonies. In other terms, they postulate a movement, i.e. a lowering, of the dominant chord in the cadence, from the original position as the head of the dominant phrase prefixed by predominant chords — 'deep structure', to the final collocation as string-adjacent to the closing tonic — 'surface structure'. At the same time, predominant harmonies do not move, so remaining in the original higher position.

In what follows, in order to facilitate the design of a categorial grammar/parser for tonal syntax, no distinction between deep and surface structure will be formulated. In the proposed template, the predominant chords prolong the full cadence, giving place to the 'prolonged cadence', rather than being the prolongation of the dominant chord, giving place to the 'prolonged dominant'.

2.1 Grammar as a Parsing System

One can refine a context-free model for tonal syntax in order to capture many subtle cases, but such models suffer an overall limitation: as traditional rule-based approaches, they can be shown to be too static and rigid; they are product-based approaches to cognition. Let me quote the following, from Lerdahl and Jackendoff (1983): 'Instead of describing the listener's real-time mental processes, we will be concerned only with the final state of his understanding'.

Lerdahl and Jackendoff introduced, as well-known, a procedure for inferring the prolongational reduction that is top-down, whose input is the time-span reduction. Such procedure is based on a set of preference rules progressively constraining the possible reductions, until the delivering of a single prolongational tree. It would be possible to formulate a bottom-up version of the prolongational tree inference, whose input is the superficial chords progression, instead of a complete time-span reduction: chords are grouped locally with their neighbours and then such small group of the lowest level are in turn clustered each other until a single, uniquely rooted tree is obtained for the entire progression. Heads of group selection, at every level, is based on a criterion of relative stability among chords. Perhaps, a similar procedure could be more psychological plausible.

Anyway, either the original top-down or the suggested bottom-up version share the same problem: they are off-line procedures. What I propose, instead, is an on-line, left-to-right procedure for deriving a plausible, not necessary the optimal, parsing tree. The procedure is also top-down because strongly biased: listeners attempt to recover a complete tonal template from the chords progression they are progressively receiving and processing. In other terms, listening to music, for listeners familiar with tonal music, is a strongly oriented process, oriented by a strong dispositional attitude. Under this respect, this process is very similar to the language understanding process, oriented by the expectation of recovering a propositional content from a string of words, given a context of utterance (Sperber and Wilson 1986, and Kempson *et al*. 2001).

2.2 The Categorial Machinery

In order to introduce a categorial machinery for tonal syntax, let me briefly introduce the fundamental template of tonal syntax (Figure 2, the right tree) in comparison with the basic normative structure in the GTTM (Figure 2, the left tree).

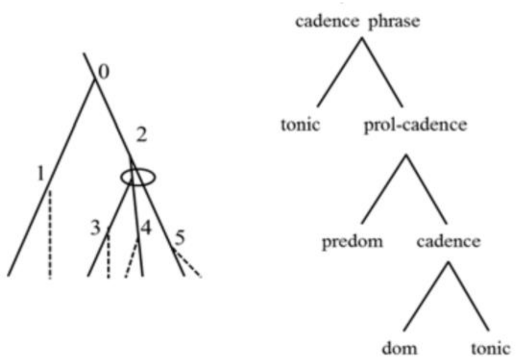

Fig. 2. The basic normative structure according to GTTM compared with the tonal template.

Categories in the template (the right tree) are to be thought of as categorization of 'prolongational regions', each region with a prolonged head, according to the following table (Table 2).

category	head
cadence phrase	prolonged cadence
prolonged cadence	cadence
predominant	*ii, IV, vi*
cadence	tonic
dominant	*V*
tonic	*I*

Table. 2. Prolonged head of each category

One can easily note the correspondence between these categories and the prolongational regions in the basic normative structure (right tree in Figure 2): node 0 dominates the region of the cadence phrase; node 1 dominates the opening tonic span; node 2 the region categorized as 'prolonged cadence'; node 3 the predominant area; node 4 the dominant area; node 5 the closing tonic. Pre-terminal categories — tonic, predominant, dominant — are directly treated as prolongational spans, without further categorization of their constituents. How to intend the heads of these categories, i.e. the scale degrees? The root of a chord basically expresses the relative stability of such chord in the tonal context defined by a key. Thus, the first degree chord (I) is the more stable chord in the key and the stability hierarchy is approximately the following:

$$I > V > IV > \{ii, iii, vi\} > vii^{o}$$

Since the root position is the most stable bass position, we can consider the heads of the categories as expressing chords preferably (not necessarily) in root position:

root position > first inversion > second inversion

We can now introduce (informally) the categorial grammar theory basic concepts (inspired by the works of Ajdukiewicz 1935; Lambek 1958). Given two grammatical categories, or types, A and B, the 'fractional type' A\B (B/A) is the one that combined on its left (right) with A outputs B:

left-concatenation:

$$\frac{A \quad A\backslash B}{B}$$

right-concatenation:

$$\frac{B/A \quad A}{B}$$

One can conceive the types above the line as the premises of a derivation of the type below the line, as suggested by the notation usually adopted in logic for representing derivations in natural deduction theory (see Van Dalen 1994). But, although the typical notation of categorial inference is in the style of natural deduction, here and henceforth we prefer the tree notation (conclusion in the root, premises in the leaves).

The first step in the formulation of a categorial grammar for tonal harmony is the choice, somewhat arbitrary, of the set of primitive types. Given the template in Figure 2 (tree on the right), we select the following categories as primitive: cadence phrase, tonic, cadence, with the abbreviations:

cadence phrase = CP
tonic = T
cadence = C

Coherently, the other categories are fractional types, defined as follows:

prolonged cadence = T\CP
predominant = (T\CP)/C
dominant = C/T

It should be remarked that this selection is mainly dictated by formal opportunities rather than by fundamental theoretical reasons (but this is a usual situation in categorial grammar theory of languages). The choice, although constrained, is conceived in order to facilitate the work of the inferential machinery (as it will be illustrated below).

2.3 The Inferential Machinery

As said, in the categorial grammar framework the characterization of syntactic structure is inferential, in the sense that a grammatical type is compositionally inferred from the types of its components. Indeed, there are three kinds of possible binary inferences, depending on what are the available premises. If the two components of a constituent are known, the constituent's type is inferred, as shown in the first line of Figure 3. Note how in the left (right) concatenation the functional type receives on the left (right) the argument of the appropriate type. This is the principal kind of inference, the so-called 'application'.

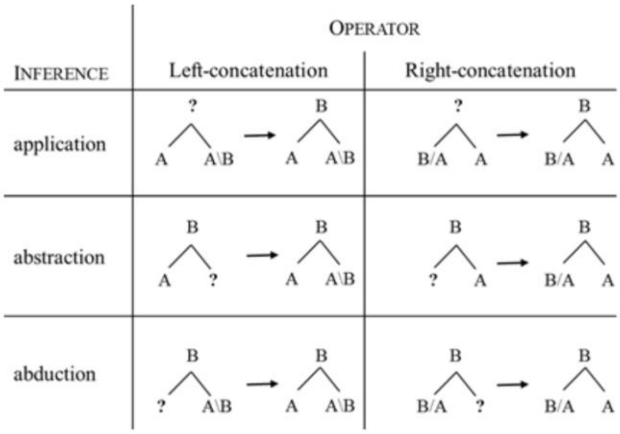

Fig. 3. The three schemes of inference in categorial grammar.

Two other inferences are possible, when the conclusion (the root of the tree) and a single premise (the argument or the functional type) are known, the so-called 'abstraction' (the second line in Figure 3) and 'abduction' (the third line in Figure 3).

Since prediction can be treated as a forward-looking inference, the inferential patterns involved in prediction are depicted in Figure 4.

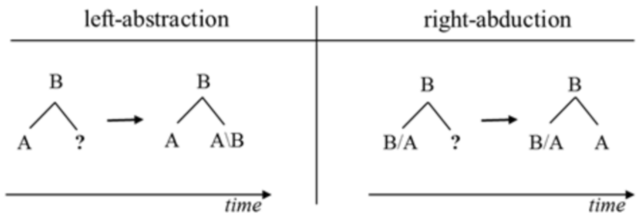

Fig. 4. The inferences for prediction.

These patterns of inferences formalize the categorial prediction in the musical parsing system. Both rules express a 'move' of prediction: given an event belonging to a category and an expected goal belonging to a category, a following event is predicted as belonging to a given category. Left-abstraction allows to predict a category of type A\B, given an input of type A and an expected goal of type B. Right-abduction allows to predict a type A, given an input of type B/A and an expected goal of type B. For example, as we will see, given a dominant chord and the expectation of the cadence, the system (i.e. the listener) predicts an incoming tonic chord.

3. THE DYNAMIC GRAMMAR OF TONAL MUSIC

3.1 The Real-Time Growth of the Parsing Tree

The dynamic grammar of tonal music (DGTM) is a set of parsing strategies for the real-time retrieval of a hierarchal structure, in the format of the basic categorial tree in Figure 4, starting from a string of chords in input. The parsing procedure begins by assuming the overarching goal of the cadence phrase

CP (default prediction)

and proceeds by alternating *scanning*, *interpretation*, and *prediction* steps, until the chord progression in input is exhausted. Before singularly illustrating these operations, some general features of the system can be summed up as follows:

– It is not necessary to provide a terminating successful state for the parsing procedure. The system terminates when the string of chord is completely scanned, with the parsing tree generated until that moment (but, as we will see, further revisions of the tree licensed, in order to satisfy the initial goal, are still possible);
– Question mark preceding a category '?X' means that X is a requirement, a prediction waiting to be satisfied. It is deleted when a chord scanned and waiting for interpretation meets the requirement (Figure 7);
– The growth of the parsing tree is triggered by left-to-right and step-by-step scanning of the input string of chords, each step releasing a transformation of the current (partial) tree (the structural context) under construction, induced by the current input;
– As already mentioned, the categorial apparatus allows to formalize prediction as inference;
– Interpretation and prediction are 'monotone' tree construction rules, devoted to the incremental growth of the parsing tree;
– During parsing, partial trees can be revised, by means of rules of tree revision (see below). These rules involve the introduction of 'non-monotonicity' in the system;
– Prediction here is equivalent to schematic expectation.

A comment is required for the first feature in this list. Differently from natural language parsing systems, DGTM is not conceived for delivering the best parsing tree, the correct, unique, structural description of a chord progression. Indeed, the system is not aimed at characterizing well-formedness of chord progressions, according to a grammar. The system is, instead, intended to formalize the listener's attempt to recover the basic tonal template from the string of chords perceived. Now, what is relevant here is not the hypothetical ending up of this attempt in a success or a failure. Rather, the system is aimed at formalizing the (cognitive) processes involved in this attempt, in the hypothesis that it can be accomplished with different degree of 'gratification' of the starting disposition. Of course, this layout raises a crucial computational issue, that of the 'termination' of the parsing process. Truly, a successful termination state can be defined as follows: the chord string in input is completely received, and the initial goal fulfilled. Consequently, one can define 'failure' when, once completed the scanning and the interpretation process, the initial goal is still waiting to be accomplished. As we will see, tree revision rules endow the system with the possibility to try to further accommodate the parsing tree, in order to obtain a more stable representation. But this is only a possibility, and one can legitimately desire that, whatever is the end-state of a parsing process, that state, as the result licensed by the system, should be intended as a model of the listener's state: so, can every description released by the system be cognitively acceptable?

In order to tackle this point, let me put the question differently. The system presupposes a 'dialectic', or flexible, tonal hearing: it is a model of a listener dynamically interacting with the musical stream, starting with a quite definite set of expectations, and 'modulating' (changing, updating) these expectations, in order to accommodate them with the items perceived. The final state of this process is not more important than the process itself, as we will see, for example, for tree revision induced by expectancy violation. However, given this dynamic interpretation of the system, it would be possible to provide it with a ‘valuator’ of the (final) representations released, in order to distinguish the more plausible ones for various typology of tonal listeners (untrained, trained, idealized).

3.2 Tree Construction

We can now proceed with the illustration of the *tree construction* operations.

Scanning. This operation is devoted to provisionally append the current chord in input at some node of the tree under construction, with a local constrain: it can be only appended at available nodes, avoiding violation of the well-formedness conditions on trees (only strict hierarchies, as shown in Figure 5). The dotted line means that the chord is in stand-by, waiting for interpretation.

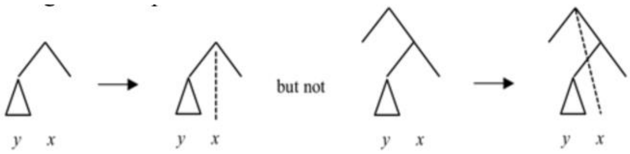

Fig. 5. The ‘scanning’ rule.

As visible in the figure, the constraint on the well-formedness of trees, applied to the current input, implicitly assures a ‘locality principle’: the scanned chord can be attached only to the left or to the right node (if available) immediately dominated by the node from which it departs. This means that a chord will be interpreted in relation to its local context.

Interpretation. This operation assigns the chord in stand-by a categorial region in its local context, as just defined (i.e. the region dominated by the node from which it departs, see Figure 6). Interpretation may be followed by *fulfilment*.

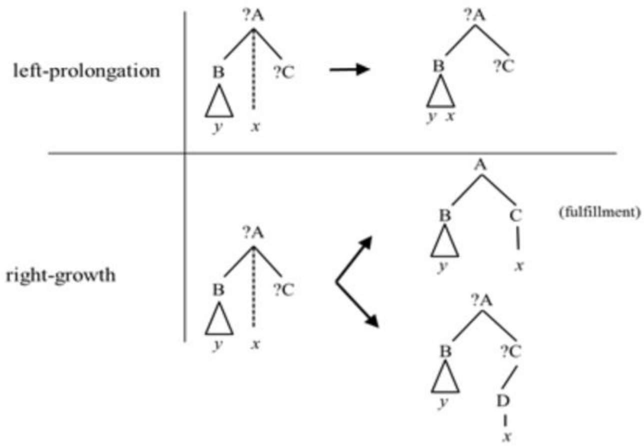

Fig. 6. The ‘interpretation’ rule.

As showed in Figure 7, there are two possible moves of interpretation of a chord in stand-by, given the structural context of the tree stored in the working memory: it can be interpreted as a prolongation of the region on its left (dominated by the higher left node departing from the node it hangs from); it can be assigned the region dominated by the higher right node, so producing a growth of the tree. If the chord satisfies the requirement (expressed by a question mark), it produces a *fulfilment*, otherwise it triggers the creation of a new node (left)-departing from the lowest requirement. The *fulfilment* of a node is inherited by its ‘ancestor’: so, if all the requirements of a node ?*x* are satisfied, then the requirement *x* is satisfied, and the question mark deleted.

Prediction. It is performed as a categorial inference. As you note in Figure 7, prediction is not triggered by a scanning step. Instead, it can be considered as a ‘spontaneous’ forward-looking inferential move, when a premise is acquired and the conclusion is expected.

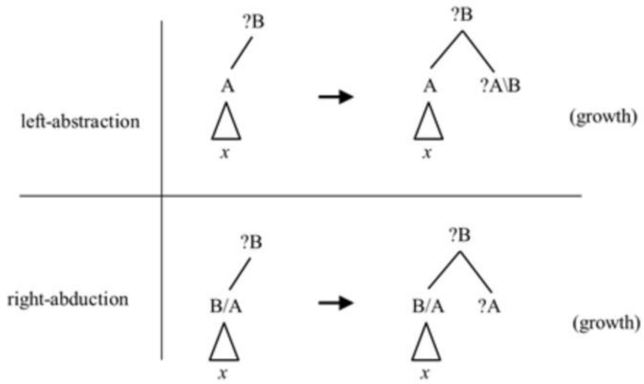

Fig. 7. The ‘prediction’ rule.

A *bias* to interpretation must be added. The initial chord is interpreted as the head of the initial tonic region (usually the I degree in root position); this is a way to initialize the system so that subsequent events are heard, and predictions are executed, relatively to such a context. As we will see, tree revision allows to modify the ‘bias expectation’, when, going on in the scanning of the input, the interpretation of the first chord as tonic ends up in an untenable, unstable, parsing tree.

3.3 A Parsing Process

In order to illustrate the tree construction mechanism at work, it can be useful to provide a simple example of the process, with some comments.

In the Figures 8–11, you can see a complete parsing procedure, in every single step. The process starts with the prediction of the tonal template. After the chord 1 is scanned, it is interpreted as tonic, by virtue of the bias (Figure 9). Then, a prediction of the prolonged cadence (T\CP) is determined, and when chord 2 is received, the system attaches it as a prolongation of chord 1 (left-prolongation). The chord 3, since it is a predominant chord, in symbol (T\CP)/C, is interpreted by the right-growth move, thus generating the prediction of a full cadence (Figure 10). The chord 4, the dominant, determines a further growth of the tree and the prediction of the closing tonic (Figure 11), in turn satisfying the starting goal (Figure 12).

The aim of this very simple example is to illustrate the joint work of interpretation and prediction in the incremental generation of the parsing tree. For example, chord 2 is attached on the left existing node of the (partial) tree, as it is felt as a prolongation of the previous event. Chord 3, instead, introduces a tension, or a distance, from the tonic area just processed, but it cannot, in itself, satisfy the expectations in stand-by (of a prolonged cadence). Thus, it is subsumed by the existing lowest (and rightmost) goal, generating a sub-goal (and producing the growth of the tree).

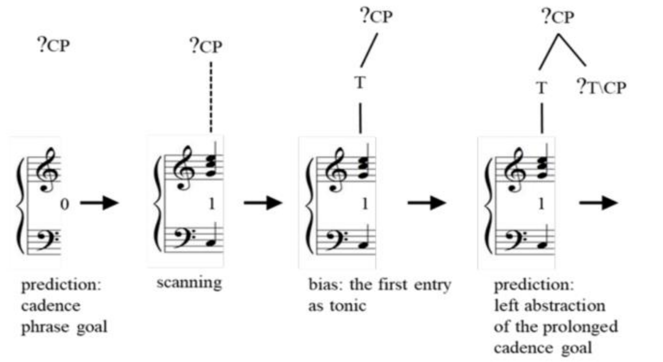

Fig. 8. A parsing example: initialization and the prediction induced by the first chord.

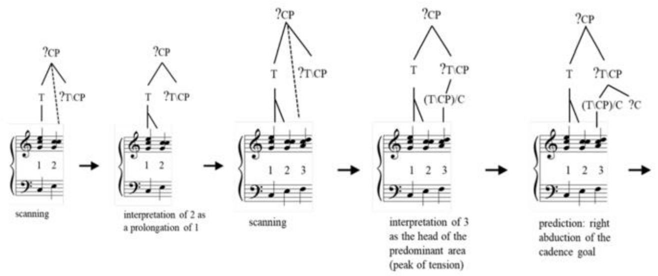

Fig. 9. A parsing example: until the prediction of the cadence.

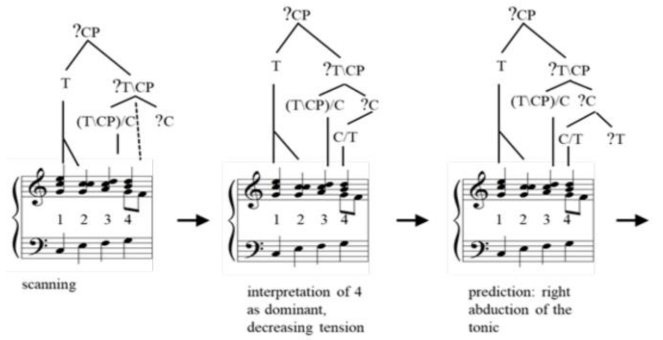

Fig. 10. A parsing example: until the prediction of the final tonic.

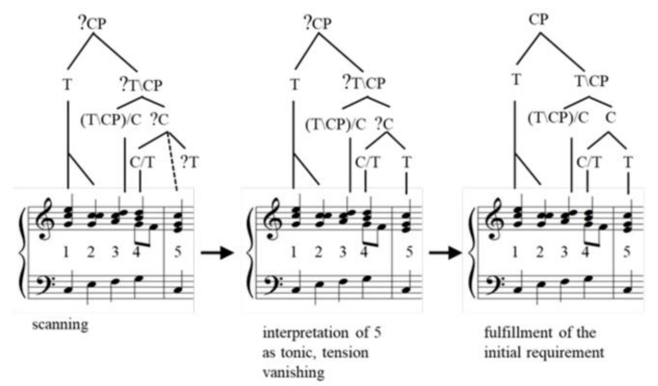

Fig. 11. A parsing example: the finalization of the process.

3.4 Tree Revision

While, as seen, the rules of tree construction are devoted to the generation of the parsing tree, the rules of *tree revision* play a completely different role in the system, thus justifying this clear distinction in the grammar. Tree revision is applied to the existing material, producing a reconfiguration of the tree. Tree revision rules do not cancel nodes of the tree they are applied, but anyway they involve the dropping of the pure 'incrementality' of tree construction.

In the essence, tree revision, in this model, is the formal device for dealing with expectancy violation. There is an important difference between language and music, relatively to expectancy violations: in language is a marked, somewhat marginal, phenomenon, specific to the performance aspects of language understanding, and forcing a re-analysis of the stream of words processed. In music, expectancy violation is emphasized and in some sense integrated in the grammar-as-parsing system. This difference specifically concerns user's performance and its pragmatic conditions, and it is relative to real-time processing of hierarchical structures, more than to structural features in themselves. We can say here that what in language is a particular phenomenon, involving a 'difficulty' in processing, in music is fully exploited as an expressive device, linking music temporal processing with emotions — this link has been mentioned in Jackendoff (1991), the only case on my knowledge.

One can mention here the so-called 'garden-path effect', widely observed in psycholinguistics (see Pickering *et al*. 2006 for a review), i.e. the difficulty in processing a sentence when a partial ambiguity rises in its real-time understanding (for example, in 'The attorney advised the defendant was guilty'). In such cases, comprehenders are forced to reanalyse the sentence to accommodate interpretation to the new material received (in the example, when they receive was guilty), i.e. to perform a revision of the current parsing. In (tonal) music, the mechanism is structurally quite similar, but with a very different semiotic function and expressive relevancy. Tree revision rules, in GDTM, are conceived to deal with this mechanism.

Structurally, tree revision rules are the device devoted to assure flexibility to the system, by allowing the manipulation of the tree under construction. A (meta)-principle of flexibility can be enunciated as follows:

Prefer the more stable partial representations.

Such principle represents the meta-rule guiding the operations of tree revision. Two general procedures are deputed to revising partial trees: *redistribution* and *projection*.

Redistribution. This rule allows the system to exchange material between existing nodes in the tree (Figure 12).

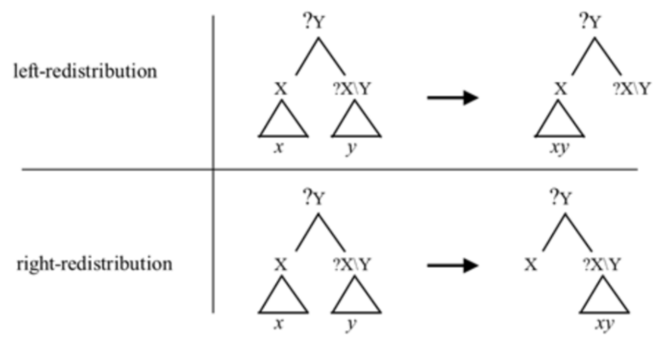

Fig. 12. Tree revision: redistribution.

Redistribution may, or may not, involve (re)-interpretation, and it can be recursively applied to its output. The only restriction is that it cannot generate new nodes, but it is possible that a node remains empty after redistribution (but it is not deleted). Left-redistribution is the really important operation, as we will see in the examples below. Right-redistribution is, nevertheless, theoretically possible, but somewhat concretely improbable. It is formulated for theoretical completeness.

Projection. It allows the system to project a new node dominating the projecting node (top-projection), or being dominated by the projecting node (down-projection). In many cases, the projected category is a cadential phrase (CP), but virtually every category could be projected (by every category, as showed in Figure 13). Top-projection is the device allowing a re-launch of the complete tonal template, when the expected conclusion of the current template is failed. This is the case of the half-cadence followed by interruption, rather than the case of the deceptive cadence. In the latter case, in fact, there is a continuation of the unfolding current template, and not a re-initialization of a new complete one — as argued for a similar context-free treatment of deceptive cadence.

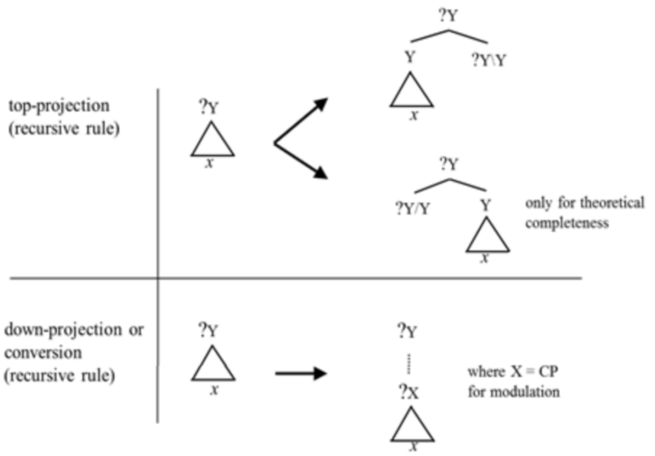

Fig. 13. Tree revision: projection.

Down-projection is devoted to create the categorial slot for modulation. In such a case, a CP is down-projected by a category that generates a temporary new tonal centre. As for right-redistribution, also right top-projection is showed (in Figure 13) for theoretically completeness.

Just to show typical cases of tree revision, the revision of interpretation is typically forced by the imperfect or deceptive cadence. In the latter case, the lack of conclusiveness of the *vi* degree chord generates the reactivation of the cadential process, as shown in Figure 14.

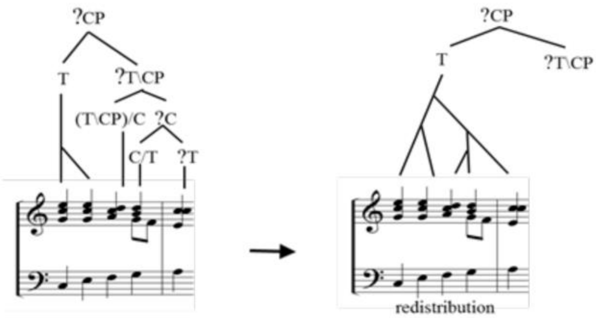

Fig. 14. An example of redistribution: the deceptive cadence.

This process can be treated as a failure of the final fulfilment, because the expected tonic is avoided. Thus, instead of the deletion of the question marks, the system predicts a new prolonged cadence. Redistribution, here, has a recursive effect.

The half-cadence and the interruption, at the end of the fourth measure, in the eight-measure classical theme in the form of 'period' (Caplin 1998) is a typical example a regeneration of expectancy, when the conclusion of a phrase is 'interrogative'. The interruption is realized by a half-cadence concluding the 'antecedent', with the V positioned immediately before its border.

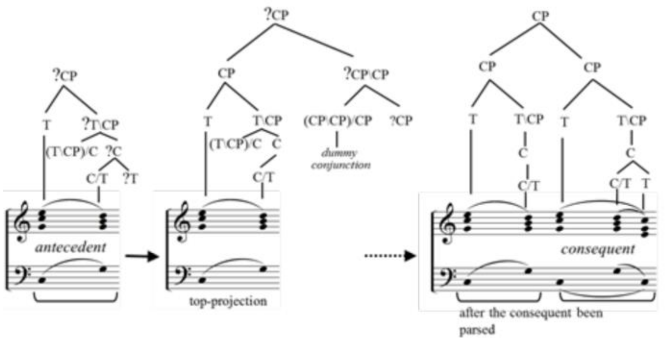

Fig. 15. An example of projection: the half-cadence at the interruption, determining the 4-measures 'antecedent' of a theme shaped as 'period'.

The break on the V, reinforced by the phrase border, induces the expectation of a following complete CP, implying a projection of a superordinate CP (Figure 15). Note the role of the 'dummy conjunction', postulated in order to reduce CP\CP to CP (in the rightmost tree).

Modulation. Adopting the well-known distinction between 'tonicization' and structural modulation, tonicization is 'absorbed' in the system as a process at the low level of pre-terminal categories' prolongation. Modulation involving temporary establishment of a different key is managed by means of down projection, according to the following principle:

Virtually every categorial region can project
a subordinate tonal centre.

Tree revision for handling modulation is forced when the default mono-tonal interpretation crashes, in order to recover affordability of the interpretation. An example is showed in Figure 16.

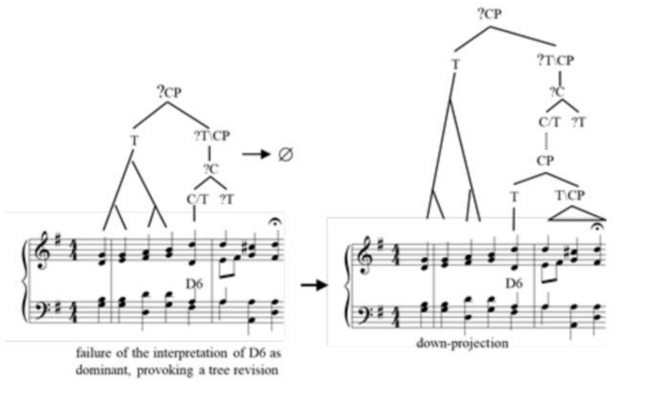

Fig. 16. Modulation as projection.

In this case (a simplified excerpt from a Bach's Choral, used also by Rohrmeier 2011), listeners cannot maintain the interpretation of D^6 as dominant. This failure forces an accommodation: D^6 is read as tonic and the remaining chords as a cadence in D major.

And what about large-scale modulation? It is at issue whether listeners have an awareness of the large-scale tonal closure (see Cook 1987). Podolak and Schmuckler (2016) have recently argued that the tonal context most influential in the perception of tones and chords is always the current key, also when it is not the main tonality of the piece — although, especially for musicians, the principal key, also in background, continues to exercise some effect. Hence, although the system could be able to release a tree for an entire piece of large dimension, one can think that the greater the dimension of the piece, the less the cognitive relevance of a unique parsing tree embracing it as a whole.

4. CONCLUDING REMARKS

Summarizing, the grammar of tonal music is a theoretical model of the real-time understanding of tonal music, built up on the presupposition that tonal understanding is strongly oriented by the expectation of the retrieval of tonal sense, and that tonal syntax is hierarchically organized. Epistemologically, this is a hypothesis of what we can call 'theoretical psychology' of music. As such, among possible desiderata, a deeper formalization could be pursued, especially for what concerns tree revision. A legitimate interrogative is whether it requires some sort of stronger constraints on its application or it can be considered a 'free procedure' of parsing tree reconfiguration, finalized at the production of more stable interpretations, when the tree under construction ends up in an untenable analysis.

Moreover, the model should be experimentally tested. Actually, what should be tested are the specific predictions of the theory regarding the different expectations generated during the listening process. Instead, the backward theoretical hypotheses are widely acknowledged in scientific literature (as already remarked above). From Meyer (1956) on, the psychological reality of expectancy in tonal listening, and its role in the emotional meaning of tonal music, is extensively observed from different perspective, until the Huron's (2006) collocation of expectancy in an evolutionary scenario. Moreover, the above discussion on hierarchies in tonal syntax was intended to show, so to say, the indispensability of a hierarchical representation of tonal structures — not forgetting the wide and enduring experimental support on hierarchies' perception in tonal cognition, from Bharucha and Krumhansl (1983). Other epistemological issues regard completeness, minimality, computational complexity, i.e. meta-theoretical properties, evidently requiring further investigations. It could be useful to remember that the model is highly 'non-deterministic': more than one output (parsing tree) is possible, and, as claimed, there isn't any ideal reading the system releases.

On the Sense of Tonal Sense

Now, the system requires further investigations. But here, in the concluding remarks, a more speculative issue can be underlined. As seen, the model postulates a tonal listening strongly oriented, the goal being the recovery of the tonal template. Then, it is natural to wonder why, as tonal listeners, we expect to recover it from a stream of musical stimuli, as a meaningful 'gestalt' of such stimuli. The familiarity with tonal idiom is certainly an important factor. Tones distributions in tonal pieces, not surprisingly, align to the interiorized Krumhansl's key-profiles (Krumhansl 1990).

But there is more. Tonal cognition is surely related to the phenomena of pitch spelling, pitch equivalence, pitch integration and grouping, melodic segmentation, and so forth. These processes can be considered as characterizing the 'low-level' processing of tonal music, and they share a lot of features with the organization of spatial-visual perception — as early noted by philosophers such as Mach and Von Ehrenfels, and more recently exploited by the seminal work of Bregman (1990). But a higher level of musical experience, more related to time-dependent dynamic profiles, seems to be equally relevant, but less investigated.

As a pure cognitive structure, the tonal template can be defined as a temporal unit, with a coherent arousal, hedonic, and emotional contour, i.e. an experience of a coherent and unitary temporal span, of an event with a given duration and a goal-oriented trajectory. So defined, the tonal template seems to be an instance of a deep 'prototype' generally involving time experience in consciousness, as articulated in temporal units. According to D. Stern (2002), this kind of organization of time cognition begins in the very initial stages of children cognitive development, playing an important role in children's development of the abilities underlying social interactions.

It can be worth quoting the Stern's notions of vitality contour and of proto-narrative envelop: 'What we mean by vitality contours are the continual shifts in arousal, activation, and hedonics occurring split-second-by-split-second that are evoked by events taking place in the body and mind of the self which are integrated into temporally contoured feelings' (Stern 1999, 70). Vitality contours are very fundamental patterns of consciousness, giving form to the experience of time, i.e. to the 'perceived present', whose length is about that of the working memory (some scholars prefer to speak of 'echoic memory'). Elsewhere, Stern (2002) introduced the 'proto-narrative envelop': 'The infant relational experiences' unit has a beginning, a middle, and an end and a line of dramatic tension. The 'proto-narrative-envelope' represents the incarnation of the internalized interactive unit. This unit is fully subjective, temporally dynamic, multi-modal, and narrative-like' (Stern 2002, 6).

Hence, the hypothesis: the tonal template could be intended as a 'cultural' projection, and sublimation of the basic unit of the subjective experience of time. Proto-narrative envelop is the developmental (and maybe evolutionary) root of the time experience as narration, and indirectly, of the tonal template. Time as narrative and the tonal template are in a strong structural or, in a sense, 'semantic' relationship: the tonal template derives its meaningfulness as a 'high-level formal pattern' of the primitive experiential consciousness of time.

Given this hypothesis, the received view according to which tonal listeners are acquainted with the tonal template because of the environmental exposition to tonal music is only 'one side of the coin', so to say, the other side being that tonal music is based on the tonal template because of its cognitive primitiveness, as sublimation of a fundamental structure (somewhat inborn) of the human experience of time.

KEYWORDS

Tonality, Expectation, Grammars, Dynamics, Parsing.

REFERENCES

Ajdukiewicz, Kazimierz, 1935. 'Die syntaktische konnexität', in Storrs McCall (ed.), *Polish Logic: 1920-1939*. Oxford: Oxford University Press, 1967, 207–31.

Bharucha, Jamshed, 1987. 'Music Cognition and Perceptual Facilitation: A Connectionist Framework', *Music Perception* 5/1: 1–30.

Bharucha, Jamshed, Curtis, Meagan, and Paroo, Kaivon, 2012. 'Musical Communication as Alignment of Brain States', in Patrick Rebuschat, Martin Rohrmeier, John A. Hawkins, and Ian Cross (eds.), *Language and Music as Cognitive Systems*. Oxford/New York: Oxford University Press.

Bharucha, Jamshed, and Krumhansl, Carol L., 1983. 'The Representation of Harmonic Structure in Music: Hierarchies of Stability as a Function of Context', *Cognition* 13: 63–102.

Bregman, Albert, 1990. *Auditory Scene Analysis*. Cambridge (MA): The MIT Press.

Caplin, William E., 1998. *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven*. Oxford/New York: Oxford University Press.

Chew, Elaine, 2014. *Mathematical and Computational Modeling of Tonality: Theory and Applications*. Berlin: Springer.

Cohn, Richard, 1997. 'Neo-Riemannian Operations, Parsimonious Trichords, and their Tonnetz Representations', *Journal of Music Theory* 41/1: 1–66.

Cook, Nicholas, 1987. 'The Perception of Large-Scale Tonal Closure', *Music Perception* 5: 197–205.

Dalen, van Dirk, 1994. *Logic and Structure*. Berlin: Springer-Verlag.

Fitch, W. Tecumseh, and Martins, Mauricio D., 2014. 'Hierarchical Processing in Music, Language, and Action: Lashley Revisited', *Annals of the New York Academy of Science* 1316: 87–104.

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

Huron, David, 2006. *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge (MA): The MIT Press.

Jackendoff, Ray, 1991. 'Musical Parsing and Musical Affects', *Music Perception* 9/2: 199–230.

Katz, Jonah, and Pesetsky, David, 2011. 'The Identity Thesis for Language and Music', <https://lingbuzz.net/lingbuzz/000959>, accessed 25/04/2023.

Kempson, Ruth, Meyer-Viol, Wilfrid, and Gabbay, Dov, 2001. *Dynamic Syntax: The Flow of Language Understanding*. Oxford: Blackwell.

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

Krumhansl, Carol L., and Kessler, Edward. J., 1982. 'Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Map of Musical Keys', *Psychological Review* 89: 334–68.

Lambek, Joachim, 1958. 'The Mathematics of Sentence Structure', *American Mathematical Monthly* 65: 154–70.

Lerdahl, Fred, 2004. *Tonal Pitch Space*. Oxford: Oxford University Press.

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

Longuet-Higgins, Christopher, H., 1987. *Mental Processes*. Cambridge (MA): The MIT Press.

Maess, Burkhard, Koelsch, Stefan, Gunter, Thomas C., and Friederici, Angela D., 2001. 'Musical Syntax is Processed in Broca's Area: An MEG Study', *Nature Neuroscience* 4/5.

Margulis, H. Elizabeth, 2005. 'A Model of Melodic Expectation', *Music Perception* 22/4: 663–714.

Meyer, Leonard, 1956. *Emotion and Meaning in Music*. Chicago (IL): The University of Chicago Press.

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

Pickering, Martin J., and van Gompel, Roger P. G., 2006. 'Syntactic Parsing', in Matthew Traxler and Morton Gernsbacher (eds.), *Handbook of. Psycholinguistics*. Amsterdam: Academic Press, 455–503.

Podolak, Olivia, and Schmuckler, Mark A., 2016. 'Understanding Modulations through Harmonic Priming', in George Vokalek (eds.), *Proceedings of the 14th International Conference on Music Perception and Cognition – ICMPC14*. San Francisco (CA):
ICMPC14. (<https://icmpc.org/icmpc14/proceedings.html>, accessed 25/04/2023.)

Rohrmeier, Martin, 2011. 'Towards a Generative Syntax of Tonal Harmony', *Journal of Mathematics and Music* 5/1: 35–53.

Rohrmeier, Martin, and Neuwirth, Markus, 2015. 'Towards a Syntax of the Classical Cadence', in Markus Neuwirth and Pieter Bergé (eds.), *What is a Cadence?*. Leuven: Leuven University Press.

Salzer, Felix, 1962. *Structural Hearing: Tonal Coherence in Music*. New York (NY): Dover Publications.

Sanguinetti, Giorgio, 2012. *The Art of Partimento: History, Theory, and Practice*. Oxford/New York: Oxford University Press.

Schenker, Heinrich, 1979. *Free Composition*, trans. and ed. Ernst Oster. New York (NY)/London: Longman. (1st ed. 1935.)

Sperber, Dan, and Wilson, Deirdre, 1986. *Relevance*. Oxford: Blackwell Publishers.

Stern, Daniel, 1999. 'Vitality Contours: The Temporal Contour of Feelings as a Basic Unit for Constructing the Infant's Social Experience', in Philippe Rochat (ed.), *Early Social Cognition: Understanding Others in the First Months of Life*. New York (NY): Psychology Press, Taylor & Francis Group.

———, 2002. *The First Relationship: Infant and Mother*. Cambridge (MA): Harvard University Press.

Temperley, David, 2007. *Music and Probability*. Cambridge (MA): The MIT Press.