Chord Context and Harmonic Function in Tonal Music

CHRISTOPHER WM. WHITE AND IAN QUINN

This article investigates several questions of harmonic function using aggressively data-driven approaches. We apply Hidden Markov Modeling—a technique used to identify contextual regularities within streams of data—to the Kostka-Payne, McGill Billboard, and Bach chorale corpora. The resulting models question the generalizability of the traditional three-function model, illustrating the syntactic uniqueness of various corpora while also highlighting recurrent characteristics of tonal repertories. Finally, this article offers some general observations, including questioning the role that tonal hierarchy plays in theories of function and discussing the cultural politics inherent in assuming the universality of one functional system.

Keywords: corpus analysis, function, tonality, Riemann, computational modeling, popular music, harmony, Kostka and Payne, McGill Billboard, Bach chorale, machine learning, music cognition, Hidden Markov models.

Since its articulation by Riemann (1893) in his *Vereinfachte Harmonielehre* of 1893, the concept of harmonic function has provided music theorists with remarkable explanatory power.[1] Function theory explains the basic intuition that certain sequences of chords are syntactical and others unexpected. What distinguishes function theory from other syntactic theories of harmony (e.g., *Stufentheorie*) is that its small number of categories yield a very efficient system of rules for harmonic progression.

But while harmonic function remains a powerful tool within contemporary music theory, with few exceptions, the concept has remained relatively unchanged since Riemann. In this article, we challenge commonly held assumptions regarding harmonic functions within music-theoretical discourse: first, while tonic and dominant are generally accepted as the basic duality of any system of functions, the number, identity, and behavior of additional functions is not necessarily clear; second, we question whether a chord's functional identity is determined by its scale-degree content, its root, its local context, or by some interaction between these factors.

Our approach is fundamentally data-driven, creating models based on large musical corpora. After reviewing the basic definitional issues concerning harmonic function, we will describe the technology used to model our data. This model will approach the idea of harmonic function from an aggressively naïve and skeptical standpoint: we will implement a model that groups chords into categories *only by the chords they tend to precede and succeed*. In doing so, we will be able to discuss certain statistical properties of the corpora themselves, thereafter relating these properties to larger discursive and theoretical issues. (We further discuss some underpinnings of our corpus methods below, along with our modeling techniques.)

We will investigate three corpora using these methods: the Kostka-Payne corpus, the McGill Billboard corpus, and the widely studied corpus of Bach chorales.[2] After describing our findings, we will return to an examination of broader topics, including some generalizations that seem to hold across our datasets that problematize music theory's standard three-function model. We will demonstrate that certain characteristics generalize across these three repertoires, but these similarities are outweighed by both stylistic differences between the repertoires and the differences in representation (e.g., Roman numerals versus lead-sheet symbols) that our models use. Our findings also connect to larger theoretical topics surrounding harmonic function: we will show evidence that surface events, such as passing tones and neighbor chords, exhibit syntactic regularities similar to consonant harmonies, and argue that the perceived hierarchical superservience of certain functions may be a simple matter of chord frequency. (Importantly, as we make clear in the ensuing prose, our goal is not to *prove* or *solve* any of these points but to provide novel observations in order to complicate and add subtleties to our understanding of harmonic function.)

CONTEXT, CONTENT, AND HARMONIC FUNCTION

We can begin to survey these issues by reviewing a frequent music theoretical argument concerning the terms "subdominant" and "predominant." The former term is often

1 The concept of grouping chords into equivalencies or hierarchies based on musical parameters precedes the formal concept of "function." Rameau (1737) groups chords into tonic, subdominant, and dominant categories based on root motions; Riepel (1755) and Momigny (1806) organize diatonic triads into a tonic-centered hierarchy; Fétis (1844) theorizes chords containing certain scale degrees as being the impetus behind harmonic tendencies; while Hauptmann (1853) generates tonal spaces using principles of dualism.

2 The Kostka-Payne corpus was used in Temperley (2009); the McGill Billboard corpus was used in Burgoyne (2012).

associated with the IV chord, though when Rameau first used the term in his *Nouveau Systême*, it referred to what we would now call ii^6_5.[3] The latter term derives from the concept of *dominant preparation* articulated by Allen Forte, under the influence of Schenker.[4] Some theorists conflate these usages: Deborah Stein, for example, writes that "the subdominant functioned either as preparation for the dominant or as a neighboring harmony that prolonged the tonic chord."[5] Other theorists seek to tease these actions apart by arguing for two distinct functions. Kevin Swinden makes the provocative claim that "harmonic function cannot be defined by pitch class alone," arguing that a IV chord should only be called "subdominant" if it progresses to I; otherwise, it is a "dominant preparation."[6] Charles Smith has gone so far as to argue that the plagal function (his term for the subdominant as distinct from predominant) should be given the same status as the dominant function, with a preplagal function analogous to the predominant.[7] The Kostka-Payne textbook argues for a still more complex understanding of IV's functional meaning: "The IV chord is an interesting chord because it has *three* common functions. In some cases, IV proceeds to a I chord. . . . More frequently, IV is linked with ii; IV can substitute for ii (going directly to V or vii°), or IV can be followed by ii (as in IV–ii–V)."[8] The authors seem to argue for a subdominant, a predominant, and a prepredominant function.

Example 1 schematizes a central aspect of the distinction between predominant and subdominant. On the left-hand side, we show a standard three-function model representing the conflation of subdominant and predominant functions. In this model, the category "S/P" represents chords including IV and other chords sharing scale degrees $\hat{4}$ and $\hat{6}$. The category is defined by its content: all IV chords belong in this category. On the right-hand side, we disambiguate the third function, showing the relationship between T and S as bidirectional, whereas a motion from T to P continues to D. Here the categories are defined by their context: whether a particular IV chord belongs in the category S or P depends on what follows it. We will make a distinction between these two ways of defining harmonic functions: *context-driven* approaches are concerned with chords' usage and, and *content-driven* approaches are concerned with chords' scale-degree constituents.

Context-driven theories define function in terms of syntax. In the right-hand side of Example 1, the difference between S and P is the *context* in which the functions respectively occur. According to Drew Nobile, "Chords gain their function by virtue of their formal position and their relationship to other chords rather than through any internal characteristics of the chords themselves. Thus, a given V chord might function as the dominant in a phrase, not because it contains the leading tone but because it resolves to the tonic and forms an authentic cadence."[9] Under this definition, functions gain their identity by their syntactic position, or the context in which the constituent chords occur.

Much of contemporary music theoretical discourse, however, focuses not on context but on content: a chord's function is determined by the scale-degree identity of its constituent scale degrees. The term *function*, of course, originates in Hugo Riemann's *Vereinfachte Harmonielehre*, which espouses an undeniably content-based approach in which a chord's function is determined by its relation to one of the three primary triads: I, IV, and V.[10] Riemann's motivation for a threefold functional system stemmed from Hegelian dialectics combined with a dualist interest in the overtone (and undertone) series.[11] But, as Eytan Agmon has argued, "the special status of root relationships by fifth is by no means a necessary assumption in the theory of harmonic function."[12] Pulling the metaphysical rug out from under function theory, some thinkers have refined their content-oriented definitions, focusing on chord prototypes rather than chord progressions, a direction Agmon himself pursues. Agmon situates the three primary triads in terms of prototype theory, where I, IV, and V are central representatives of three categories maximally distant from one another in a voice-leading space (see Ex. 2). He does not, however, question Riemann's fundamental assumptions about the number and identity of the harmonic functions.

Contemporary theorists seeking to extend the concept of harmonic function beyond the Riemannian limits invoke the expectations, qualities, or hierarchical positions associated with specific scale degrees. Daniel Harrison retains a threefold functional system but considers individual scale degrees as the source of functional identity, with scale-degree assemblies rather than rooted tertian chords as the bearers of harmonic function.[13] Fred Lerdahl similarly derives function from the placement of pitches and chords within a tonal hierarchy, with tonic as the most superordinate, dominant the next, and so on.[14] To the three Riemannian functions, Lerdahl adds an additional four (departure, return, neighbor, and passing) based on the distinct ways that subordinate chords can prolong

3 Lester (1992, 132).

4 Forte (1962). The Schenkerian concept of intermediate harmony includes a wider range of chords than we traditionally assign to the subdominant and predominant categories.

5 Stein (1983, 156).

6 Swinden (2005, 253).

7 Smith (1981). Tymoczko (2003) notes that the concept of plagal function might find its roots in the Boulanger tradition.

8 Kostka and Payne (2012, 114), emphasis added.

9 Nobile (2014, 22).

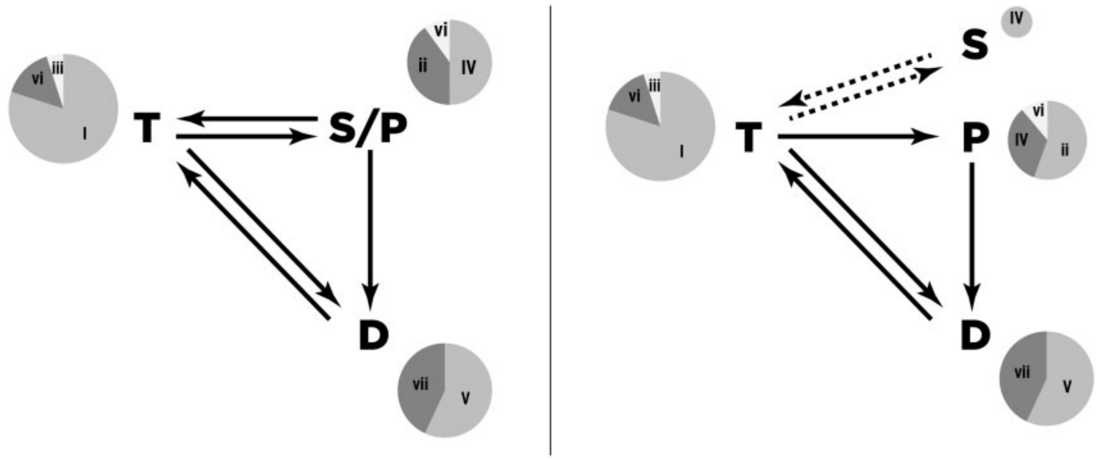
10 Harrison (1994, 39–40); Kopp (2002, 5).

11 "These ist die erste Tonika, Antithese die Unterdominante mit dem Quartsextakkord der Tonika, Synthese die Oberdominante mit dem schliessenden Grundakkord der Tonika; thetisch ist die Tonika, antithetisch die Unter-, synthetisch die Oberdominante" (Riemann [1900–1901, 3]). For more on Riemann's dialectics, see Harrison (1994), Klumpenhouwer (2007), and Dahlhaus (1968 [1990]).

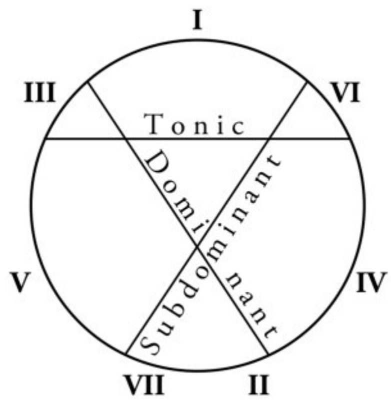
12 Agmon (1995, 199); see also Agmon (2013).

13 Harrison (1994).

14 Lerdahl (2001).



EXAMPLE 1. *Two probabilistic models of harmonic function*



EXAMPLE 2. *Agmon's Circle of Functional Prototypes*

relatively superordinate chords. Lerdahl's *Dep* function, for example, characterizes chords that, while lacking a leading tone, otherwise resemble back-relating dominants in having a temporal and prolongational dependency on a preceding tonic. In one sense, this function abstracts away from the specific scale-degree constituents of dominant harmony, but, as with each of Lerdahl's prolongational functions, its identity is determined wholly by a prolongational structure that is itself dependent on a tonal hierarchy. What seems like a context-oriented function is, in fact, tethered to an elaborate and precisely determined notion of a hierarchical pitch space. What remains unexplained, however, is how scale degrees are endowed with this hierarchical power.

While most investigations of function start by intuiting that certain groups of chords progress to other groups of chords and build functional categories from that standpoint, this study begins from the intuition that music simply has contextual regularity and then investigates one way in which chords might group into those contextual categories. The power of this approach is threefold: first, it allows us to isolate one particular corner of harmonic function—two-chord context—and isolate the behavior of groups based exclusively on that parameter; second, it allows us to test our intuitions about chord content; third, our method allows us to ask, "what about this corpus invites us to hear syntactic categories?"—allowing us to begin to connect our computational findings to our intuitions surrounding harmonic function. To these ends, we propose a model that takes into account *context only*, ignoring the scale-degree content of chords. The approach is based on Hidden Markov models (HMMs), a technique from machine learning described in the next section. Our project is aggressively data-driven: our results will model the contextual properties of large amounts of musical data in a way that attempts to minimize the effects of preexisting theoretical assumptions on the analysis. Our goal in taking a radically data-driven approach is not to create a new theory from scratch or to discard existing theories, but to isolate one parameter of harmonic function (chord context) in its most simple form (two-chord progressions) in order to both learn about the syntactic properties of the corpora in question and to interrogate our own ideas about harmonic function.

SOME NOTES ON CORPUS METHODOLOGY

Given that corpus methods are relatively novel within mainstream music theory, a few words on our motivations and methodologies are in order. Digital humanities scholars frame their computational analyses in a number of ways, but, for our purposes, we might divide approaches into two rough categories, *the backhoe* and *the microscope*.[15] "Backhoe" methods use computers to do the sorts of things that humans do, just faster. From this standpoint, the analyses that follow represent insights into chord function that one might come to given the painstaking chord-to-chord annotations and repeatedly calculating groupings based on those similarities. "Microscope" methods suggest that their computational analyses illuminate

15 These categories are roughly drawn from Wittig (1977).

things that we could not see with purely human capacities. From this standpoint, our analyses offer observations into the way that chord-to-chord progressions might group together, because they do so with a precision, insight, or impartiality impossible to human perceptions.

The ability to strictly isolate the contextual parameter in order to create broader contextual categories is arguably unique to the domain of computation, and therefore more of a "microscope" approach. But, regardless of how exactly our approach situates within this continuum, it is important to keep in mind what the method does *not* do. Regardless of the formalisms, our goal is not to "discover" any sort of "truth" underlying harmonic function—our models are neither meant to be strictly predictive of human perception or analysis, nor to say what theorists or musicians should perceive or analyze, nor even to capture the compositional models used to produce these repertoires. Rather, we will investigate the notion of function by isolating one aspect of its definition—context, which in its simplest form entails two-chord progressions—by producing a formal model using only those parameters.[16]

But, while our methods are not meant to strictly model either the methods of composers or the behavior of listeners, they do allow for speculative connections into these domains. After all, corpora were produced by composers and designed to be heard and understood by some group of listeners. As described in White (2013), while not producing definitive evidence, corpus studies allow us to discuss the properties that are observable in some composer's style, as well as the kinds of statistical regularities that allow listeners to both learn about and interpret music.[17]

However, unlike other music-theoretic methodologies, the adoption of a computational approach compels researchers to define precisely how their data will be collected, modeled, and interpreted. A microscope will always produce two-dimensional images of the cells and microorganisms it observes, but that does not mean we believe that these images prove that cells and microorganisms are two-dimensional creatures. In the project documented in this article, our adoption of HMMs produces results that tell us both about contextual regularities in musical corpora and about the predilections of HMMs. It is left to us to interpret the results in a way that foregrounds the former and domesticates the latter.[18] Other computational models of harmonic function will produce different results, given their contrasting inner workings. In what follows, we will endeavor to make clear which features of our results are more interesting for music theorists and which are side effects of our reliance on HMMs.

COMPOSITE HIDDEN MARKOV MODELS

Example 1 uses three kinds of symbols: Roman numerals signifying chords, letters signifying functions, and arrows signifying motions between functions. The logic of these models involves an underlying progression of functions and a system for realizing those functions as chords. A given functional progression can be realized in multiple ways: T–P–D–T, for example, might be realized as I–IV–V–vi or as iii–vi–V–I. In the language of Hidden Markov Models, chords are *observations* generated by underlying *hidden states*, or functions.[19] The basic assumption of the HMM formalism is that a given observable sequence is determined by two sets of probabilities: *lexical probabilities* determine what observable form will be taken by any given hidden state, and *transition probabilities* determine the progression of hidden states. Example 3 illustrates the generative idea of HMMs with its characteristic interaction of these two types of probability. The likelihood that one observed symbol will follow another is determined only indirectly with hidden states governing both order (transition probabilities) and observed form (lexical probabilities). A HMM is therefore comprised of four items 1) a set of *k* possible hidden states, 2) a vocabulary of possible observed symbols, 3) a probability distribution that determines how often each hidden state proceeds to each other hidden state (transition probabilities), and 4) a probability distribution governing how likely each vocabulary item is to be produced by these hidden states (lexical probabilities).

The transition probabilities represented in Example 1 (the arrows) predict that T–P–D–T is a much more likely functional progression than D–P–T, and the lexical probabilities (the pie charts) make it unlikely that T–P–D–T would be realized as IV–iii–I–V.[20] Again, the model does not directly predict what chords should follow a given chord. There is no rule that directly prohibits ii from moving to I, but the combination of transition and lexical probabilities make this succession highly unlikely. The HMM formalism does not depend on

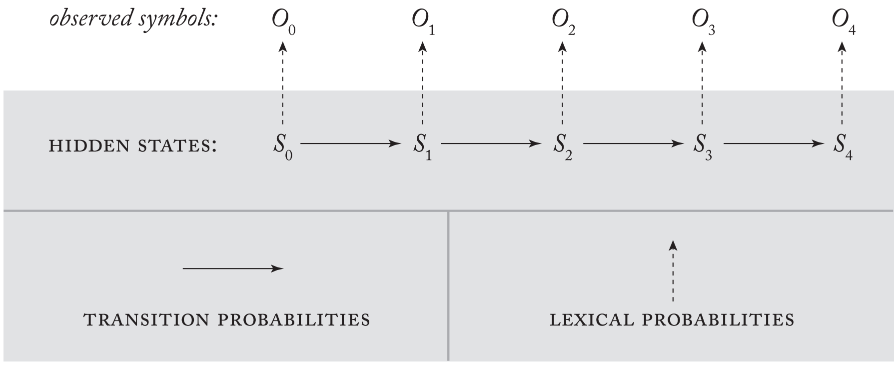
16 This somewhat differs from how statistical modeling is used in scientific inquiries of music (e.g., Krumhansl 1990, Temperley 2009, Burgoyne, Wild, and Fujinaga 2013) but adheres to norms of digital humanities. While in the former, computational models can be judged by their adherence to a ground truth ("correct answers") of human annotations or responses, the latter observes some trend within a dataset in order to deepen our understanding of the dataset itself.

17 White (2013). The latter claim is based in theories of statistical musical learning in which humans learn musical norms by being exposed to the statistical properties of some dataset (e.g., Saffran, Johnson, Aslin, and Newport 1999; Creel, Newport, and Aslin 2004; Saffran, Reeck, Niebuhr, and Wilson 2005; Huron 2006; Loui, Wessel, and Hudson Kam 2010; Loui 2012).

18 Mavromatis (2009, 2012) provide an excellent case study of this kind of interpretation.

19 HMMs, while used frequently in linguistic studies, are relatively new to music theory research. Panayotis Mavromatis (2005) has experimented with HMMs to investigate meter in Palestrina and the syntax of Greek Orthodox chant, while several researchers (e.g., Bello and Pickens [2005]) have used HMMs to induce harmonic content from sound waves. Raphael and Stoddard (2004) have used HMMs to fit pitch-class information into mode, key, and chord templates.

20 The absence of arrows between two states most often indicates that the transition is highly improbable rather than impossible.



EXAMPLE 3. *The generative model of HMMs*

prototype theory, though lexical probabilities can strongly favor a particular realization of a given hidden state. For instance, in the case of T, the most likely observed chord is I. (At this point, it is important to remember that, even though we might use symbols that denote chord content to a human, these characteristics are unavailable to the underlying algorithm. To the computer, "I" and "vi" are arbitrary symbols, and it has no way to know that chords designated by these symbols share two common tones. Rather, the algorithm is only aware of how these random symbols are positioned in relation to one another in the analyzed corpus. It is in this sense that our model is "purely contextual.")

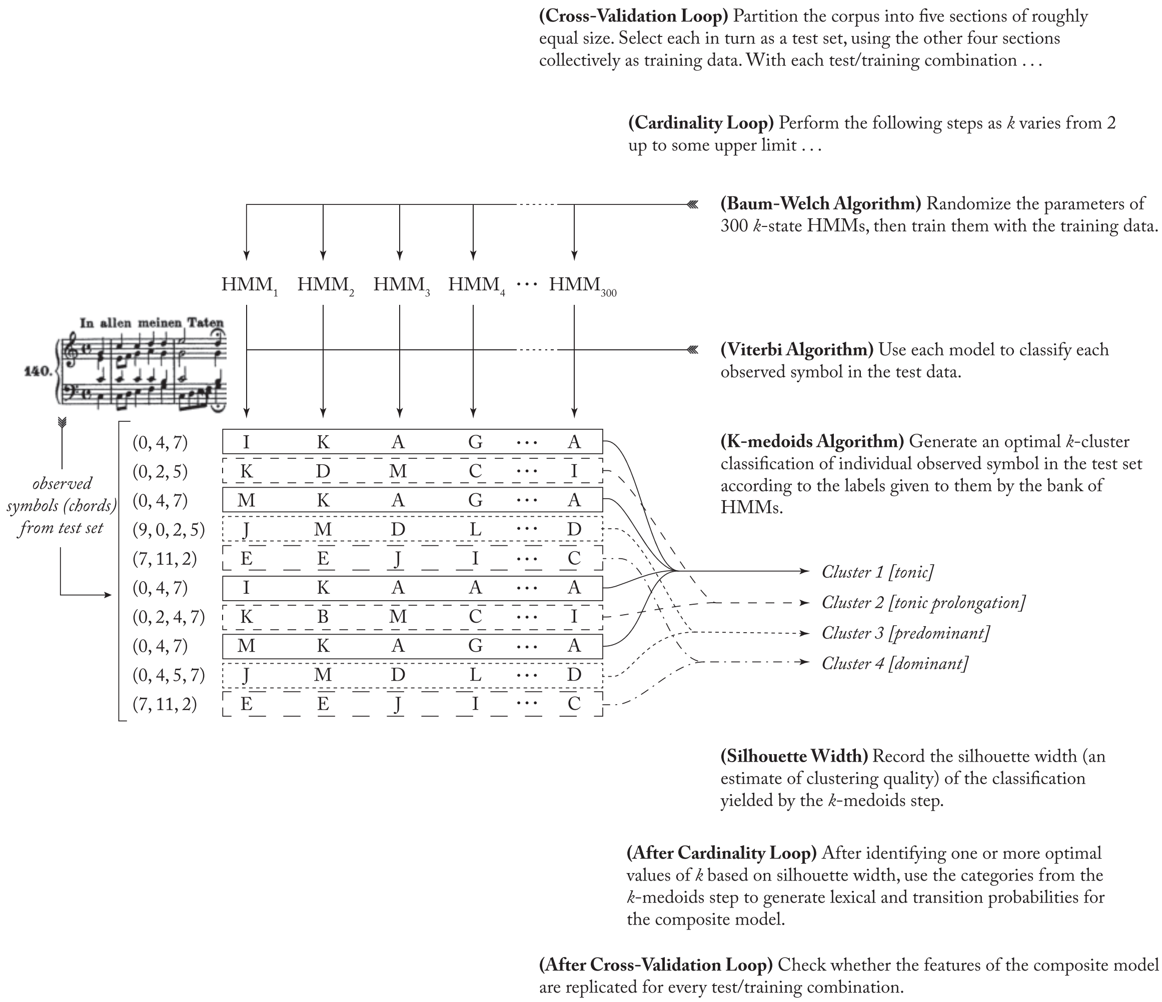
If lexical and transition probabilities are known, the model can be used to analyze a series of observations, assigning the most likely hidden state to each chord. This task is comparable to what we teach our undergraduate theory students to do when undertaking functional analysis: to reconcile an observed musical surface with an abstract knowledge of the rules of functional progression. For example, knowing whether to call a given vi chord tonic or predominant depends both on what the next chord is and what the acceptable functional progressions are.

Techniques of machine learning make it possible to estimate transition and lexical probabilities directly from a given series of observations. An iterative process called the Baum-Welch algorithm finds an optimal fit between the hidden states and the observations. Baum-Welch begins with a random set of transition and lexical probabilities, and each iteration of the algorithm consists of two steps known as *expectation* and *maximization*. In the *expectation step*, these (initially random) probabilities are used to *decode* or analyze the observation sequence, determining the most likely hidden state for each observation given the current model. The model can then assign an overall probability to the sequence. Initially this probability will be low, since the model's parameters (the transition and lexical probabilities) are set randomly. In the *maximization step*, the algorithm adjusts the parameters in an attempt to improve the probability of the training data given the model. The expectation-maximization process is iterated until the improvements reach a point of diminishing return. A more detailed discussion of HMMs and examples of expectation maximization are presented in Appendix A.[21]

While the Baum-Welch algorithm runs largely unsupervised, it requires an important decision to be made in advance: the number of hidden states (k). While there is no universally accepted method to determine the ideal number of states, we introduce a novel approach in this essay. Our method exploits a property of expectation-maximization algorithms, like Baum-Welch: in part because of the random initialization of the parameters, models estimated from the same training data in repeated trials will not replicate each other consistently if the chosen value of k does not fit the data well. Example 4 depicts the procedure; we will begin our discussion *in medias res*, at what is labeled as the Cardinality Loop. For each value of k under consideration, we use Baum-Welch to train a bank of 300 HMMs on the same training data. We then use a standard method (the Viterbi algorithm) to have each HMM decode a *test data* that was not included in the training data. Every observation (chord) in the test data is then associated with a vector of 300 hidden-state labels, each assigned by one of the HMMs.

The novelty of our procedure lies in the next step, where we attempt to create a composite of the 300 HMMs. We use a different expectation-maximization algorithm, known as *k-medoids*, to classify the set of observations in the test data into k categories (where k is the number of hidden states in each HMM) based on a simple dissimilarity measure equal to the square of the number of HMMs that decoded the two observations in question into the same hidden state. The k-medoids algorithm, as it is implemented in the R programming language, yields a quantity called *silhouette width*, a measure of the ease of the algorithm's task given the data. A low silhouette width indicates overlapping and interpenetrating categories. The higher the silhouette width, the clearer the boundaries between categories are, and the higher the degree of consistency between the individual HMMs. A high silhouette width allows us to treat the categories learned by the k-medoids algorithm as hidden states of a composite of all the HMMs.

21 See the online version of this article.



EXAMPLE 4. *Procedure for learning and cross-validating composite HMMs with optimal numbers of hidden states*

We prefer values of *k* for which the HMMs replicate each other more consistently (indicated a higher silhouette width) than for neighboring values of *k*. A higher silhouette width indicates more consistent solutions within the hundreds of models using that *k* value: if that value produces a peak compared to the surrounding values, that means *k* hidden states organize the space better than both k-1 functions, and $k+1$ functions. In this study, we prefer values of *k* exhibiting the most dramatic peaks. All silhouette-width figures are shown in Appendix A.[22]

There are two approaches to validating an HMM. One way is to use it to decode some test data and compare the results to a *ground truth* derived from human analyses of the same data. If the model reproduces the insights of the human analyst, the correlation suggests that the model is producing some salient result. We are, however, less interested in how a modern listener might analyze a corpus than we are in identifying consistent and reproducible properties of corpora. To this end, we also use *cross validation*, in which we create a model using only some fraction of the dataset and then assess how well the model conforms to the remaining fraction. If one iterates this process with different divisions and the same results hold, one can be relatively certain that the model is capturing some generalizable and robust property (rather an artifact of random variation). The following analyses vary the types of validation used: since the Kosta-Payne corpus is derived from a textbook that addresses issues of harmonic function, we compare our results to the authors' own statements on this topic. Given that we mean to be skeptical of how notions of harmonic function manifest on musical surfaces, the Bach chorale and popular music corpora are both sufficiently large to engage in the latter cross-validation tactics.

Such a model of harmonic function takes only the context of chords—and not their content—into consideration. If the model puts two chords into the same functional category, it is on the basis of their behavior alone and not any shared scale-degree content. In what follows, we apply this modeling procedure to several corpora, beginning with a relatively straightforward dataset: the Kostka-Payne corpus.

22 See online version of this article.

MODELING THE KOSTKA-PAYNE CORPUS WITH FOUR FUNCTIONS

The Kostka-Payne corpus comprises the analytic annotations to the musical examples provided in the instructors' edition of the eponymous harmony textbook. This corpus records the root of each chord in both absolute (pitch) and relative (scale degree) terms, along with metadata about the composer, title, and mode (major or minor) of each excerpt. Relative-pitch chord roots are represented as integers modulo 12, corresponding to chromatic scale degrees: I = 0, ♭II (or ♯I) = 1, II = 2, etc. These representations are equivalent to Roman numeral analyses that specify neither inversions nor the existence of nonchord tones or chordal sevenths, with the qualification that no enharmonic distinctions are made between, say, ♭II and vii°/ii, or ♭VI and vii°/vi. Chord quality is also not included in these annotations: V/ii and vi, for instance, would receive the same annotation since they share the same root. Temperley provides two versions of the corpus that differ in their treatment of cadential $^6_4$ chords: we use the version in which these chords are assigned a root of $\hat{5}$. Example 5 shows the transitions between the chords within the major-mode segment of the corpus, with the most frequent transitions in bold (transitions are from rows to columns).

The Kostka-Payne corpus provides an ideal starting point for this project, since we can compare the authors' own discussion of each chord's contextual role with our results. We, therefore, began our HMM analysis of this corpus with a series of three-state models, to see whether the HMM would sort chords into the conventional tonic, dominant, and sub/predominant functions. The clusters at $k = 3$ produced a low silhouette width, indicating a failure of the individual HMMs to converge collectively on a single composite model. Example 6 shows two sample solutions produced by the EM algorithm. Transition probabilities are represented by the weight of the arrows, and lexical probabilities are represented by pie charts corresponding to the hidden states. Due to the corpus's ambiguity regarding chord quality, chord roots are represented as scale degrees in Arabic numerals rather than Roman numerals. The size of each pie chart represents the relative probability of each hidden state. Specific probability tables can be found in Appendix B.[23]

Many aspects of the model in Example 6(a) resemble the idealized model in Example 1. Moving clockwise from the upper left, the three states correspond to tonic, predominant, and dominant functions, and the chord roots output by each hidden state are those predicted by traditional three-function theory. This model, however, is not representative of the family of models produced by the Baum-Welch algorithm. Since its transition probabilities are distributed relatively evenly—no pathway is unidirectional and no arrow is particularly thick—the model's probability judgments will be somewhat low. Example 6(b), on the other hand, contains several unidirectional arrows, but its lexical probabilities are more evenly distributed. In particular, to make sense of IV–I progressions, the model places IV chords into the "dominant" category, essentially conflating S and D functions under Example 1(b) into a single "pretonic" category. Chords with a root of $\hat{6}$ have also been removed from the tonic category in this version: while this simplifies the lexical probabilities by placing all such chords into the predominant category, it makes deceptive cadences very unlikely events. Because of these difficulties, this model also assigns relatively low probabilities to the observation sequences. Other three-state models are similarly compromised, though in different ways, and the composite three-state model is incoherent, suggesting that this value of k does not fit the contextual regularities of the Kostka-Payne corpus.

When $k = 4$, the Baum-Welch algorithm produces a family of models that are more consistent with each other. Example 7 shows the composite solution, again using arrows for the transition probabilities and pie graphs of Arabic numerals for the lexical probabilities. Moving clockwise from the upper left, these hidden states function as tonic, pre-predominant, predominant, and dominant/pretonic.[24] (In what follows, we will refer to these as **T**, **P⁻**, **P**, and **D/T⁻**.) The most frequent transitions occur between **T** and **D/T⁻**, with more transitions from **D/T⁻** to **T** than vice versa. **T** goes to several functions, but **T** is only progressed to via the **D/T⁻** function. The remaining two functions offer two other routes to prolong the **T–D/T⁻–T** progression, either first visiting the **P** function before going to the dominant, or prepending the **P⁻** into that progression. Notably, this four-function diagram reflects Kostka and Payne's analysis (discussed above) of the "three common functions" of a IV chord: proceeding to I, proceeding to V, or proceeding to ii, which will in turn proceed to V. This diagram also illustrates the authors' preference for falling fifths and thirds within chord syntax.[25] The most probable chord progression involving all four functions, given these lexical and transition probabilities, would be I–vi–ii–V–I, a prototypical falling fifth/third progression.

Importantly, these results conform to the authors' description of chord function. We quoted above a crucial passage concerning the IV chord, suggesting three functions for that sonority: in the composite four-function model, IV chords do indeed have three possible roles. Furthermore, our HMM's behavior also reflects many aspects of Kostka-Payne's chord-function diagram, reproduced below in Example 8. Just as in the four-function HMM, the diagram shows a "dominant" category with V and vii° progressing to I, a "predominant" category with ii and IV progressing to the dominant chords. Unlike our model, Kostka and Payne group iii and vi chords into their own categories; but like our model, these chords progress between one another and progress to the predominant category. While the iii and vi chords may have their own independent transitional tendencies (iii, for instance transitions to vi most frequently, while vi transitions most often to ii), the

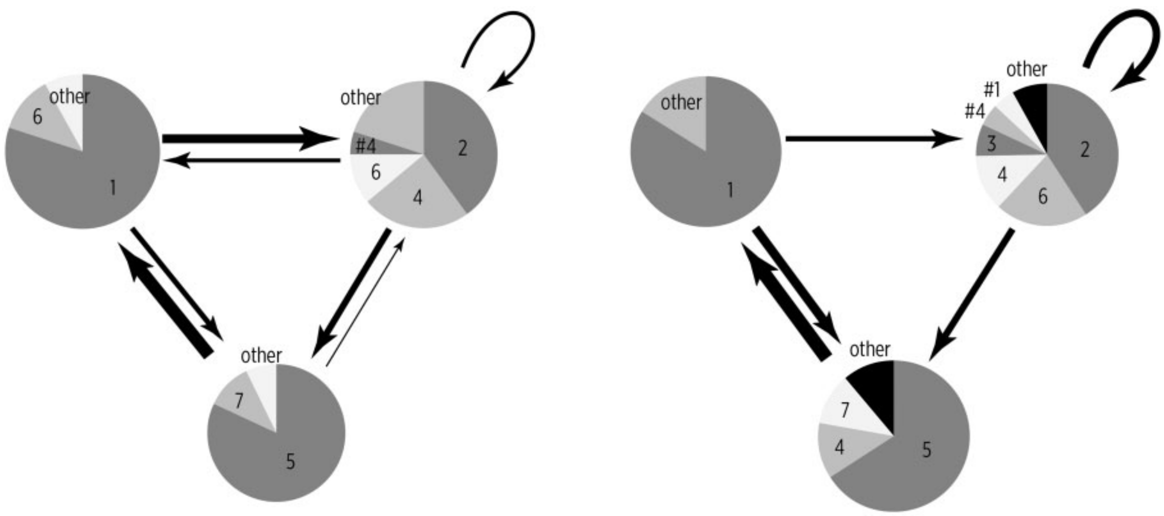
23 See online version of this article.

24 "Pre-predominant" is adopted from Doll (2007).

25 Kostka and Payne (2012, 104–5).

	I	vii°/ii	ii	♭III	iii	IV	vii°/V	V	bVI	vi	vii°	N/A
I		4	24	1	5	32	1	51	5	10	11	6
vii°/ii	1		8							1		
ii	4	3			4	1	6	49	2	10	6	1
bIII								1	1			
iii	1		2			5		1		7	1	
IV	21	1	7		3		3	10		1	3	
vii°/V	3					1		8				
V	116		7		2	3			3	7	2	5
♭VI	3		2					3		4		1
vi	4	2	29		1	5	2	1			1	2
vii°	20			1	3			2	2			
N/A	3		1					11				

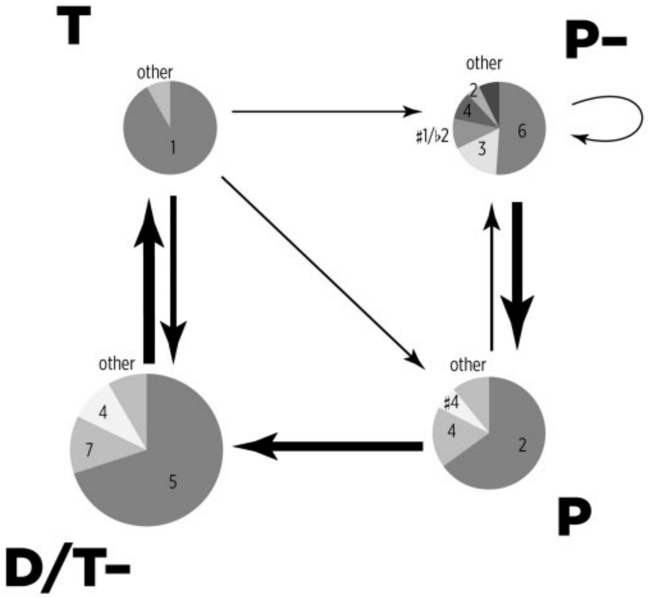
EXAMPLE 5. *Transitions between chords in the major-mode segment of the Kostka-Payne corpus*



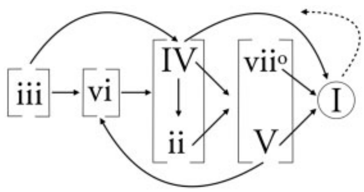
EXAMPLE 6. *Two representative three-state models: (a) and (b)*

overall interactions of these behaviors create the contextual categories captured in this model.[26]

26 This feature of the resulting models makes them often insufficient to generate actual chord progressions. For example, a vii°/ii chord should generally proceed to a ii chord in actual composition; however, the HMM would seem to indicate that this chord could proceed to any other chord within the P- or P functions.



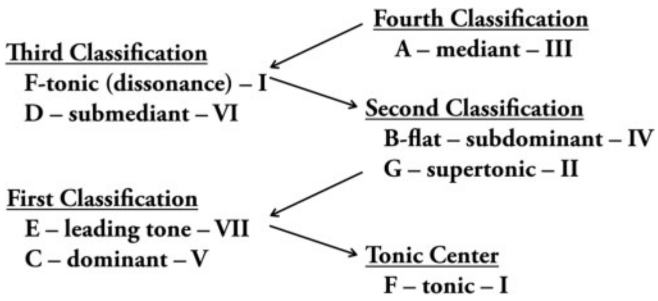
EXAMPLE 7. *The composite solution for four states*



EXAMPLE 8. *Kostka and Payne's chord-progression chart*

But this model also reflects the outlook of other work linking harmonic function to corpus analysis. Allen Irvine McHose, an early advocate of a corpus-based approach to tonal harmony, uses a model strikingly similar to our four-state model in his 1947 textbook, *The Contrapuntal-Harmonic Technique of the 18th Century*. Undertaking an exhaustive quantification of the root motions in "Bach, Graun, Handel, and Other Contemporaries," McHose groups chords into four "classifications," shown in Example 9.[27] The classifications schematize a series of falling fifths into a final tonic, creating five equivalence classes reminiscent of the four-function model under Example 7.

The most peculiar property of our model, which sets it apart from McHose and nearly every other textbook, is that of "dominant" IV chords. The model conflates dominant and subdominant functions, creating a "pretonic" hidden state. This conflation is a consequence of the model's exclusive focus on two-chord context: without knowledge of scale-degree content or common tones, the model unifies the dominant and subdominant functions, recognizing that they tend to occur in the same contexts. (Notably, $k = 5$ solutions neither



EXAMPLE 9. *McHose's Four-Chord Classifications*

disambiguate these categories, nor produce consistent results.) (See the discussion of Examples A4–6 in Appendix A.)

Our analysis of the Kostka-Payne corpus, then, has highlighted two issues: the surprising failure of the Baum-Welch algorithm to learn the standard three-function model from the corpus, and the inability of a purely contextual model to segregate chords in some ways valued by traditional function theory, typified by "dominant" IV chords. In what follows, we expand our investigation into pop and rock music, a corpus in which the IV chord has an even wider array of uses. This move allows us not only to access one of the largest publicly available corpora of Roman numerals, but also to engage directly with the in-depth, critical, and contentious scholarly engagements with harmonic function that can be found in the literature on pop-music analysis.

MODELING THE MCGILL BILLBOARD CORPUS WITH EIGHT FUNCTIONS

Scholars disagree about some very basic definitions about harmonic function in popular music, ranging from what constitutes dominant function to whether traditional harmonic function has any relevance to pop-music at all. On the one hand, scholars like Allan Moore, Ken Stephenson, David Temperley, and Trevor deClercq argue for a fundamental difference between popular music and common-practice harmony.[28] Temperley and DeClercq, in a corpus study of rock harmony, emphasize many aspects of this difference: rock harmony does not have strong unidirectional tendencies (e.g., V progresses to IV as much as IV progresses to V), and, in many cases, IV (rather than V) functions as the primary non-tonic triad. On the other hand, several analysts have attempted to theorize pop/rock harmonic syntax as an extension of common-practice norms. Nicole Biamonte and Chris Doll, for instance, argue for including modal harmonies into functional models, with ♭VII functioning as dominant (Doll's "rogue dominant") or as IV/IV (Biamonte's "Double Plagal" progression).[29] Going even further, Drew Nobile entirely dissociates traditional harmonic functions from the scale-degree content

27 McHose (1947, 4–9).

28 Moore (1992); Stephenson (2002); deClercq and Temperley (2011).

29 Biamonte (2010), Doll (2007).

of chords.[30] In Nobile's formalization, almost any chord can function as a tonic, dominant, or predominant: "a chord's function is given more by formal considerations—i.e., what role it plays within the form—than by its internal structure or any specific voice-leading motion."[31] Nobile allows for predominant V chords, dominant IV chords, and so on.

MATERIALS AND METHODS FOR THE MCGILL BILLBOARD CORPUS

For this study we performed an HMM analysis on the McGill Billboard Project dataset, one of the largest corpora of Roman numeral progressions publicly available.[32] This dataset includes harmonic annotations for 649 randomly chosen songs appearing on Billboard's "Hot 100" list between 1958 and 1991. The annotations were made by human analysts and included information about root and chord quality. The entire corpus consists of 270,366 chords. We removed annotations that duplicated the immediately previous annotation, leaving 64,591 chords. The corpus's key annotations were used to assign scale degrees: given that mode is not designated in this corpus's annotations and the major/minor duality is not as clear in popular music as in classical music,[33] we elected to use all pieces regardless of mode. Stretches of chords in a single key were used as the observations for the HMMs.

To ensure the statistical robustness of our models, we followed a *cross-validation* procedure: we divided the corpus into five equal parts, training each HMM on four parts (the *training set*) and using it to analyze the remaining part (the *test set*). For each value of *k* between 2 and 30 (inclusive) and each test set, we produced 300 models.[34] As with our procedure for the Kostka-Payne corpus, we optimized *k* by identifying the most consistent group of models with the added constraint of seeking agreement among the five test sets. The most consistent solution involved eight states, shown in Example 10. The specific probability tables can be found in Appendix B.[35]

RESULTS

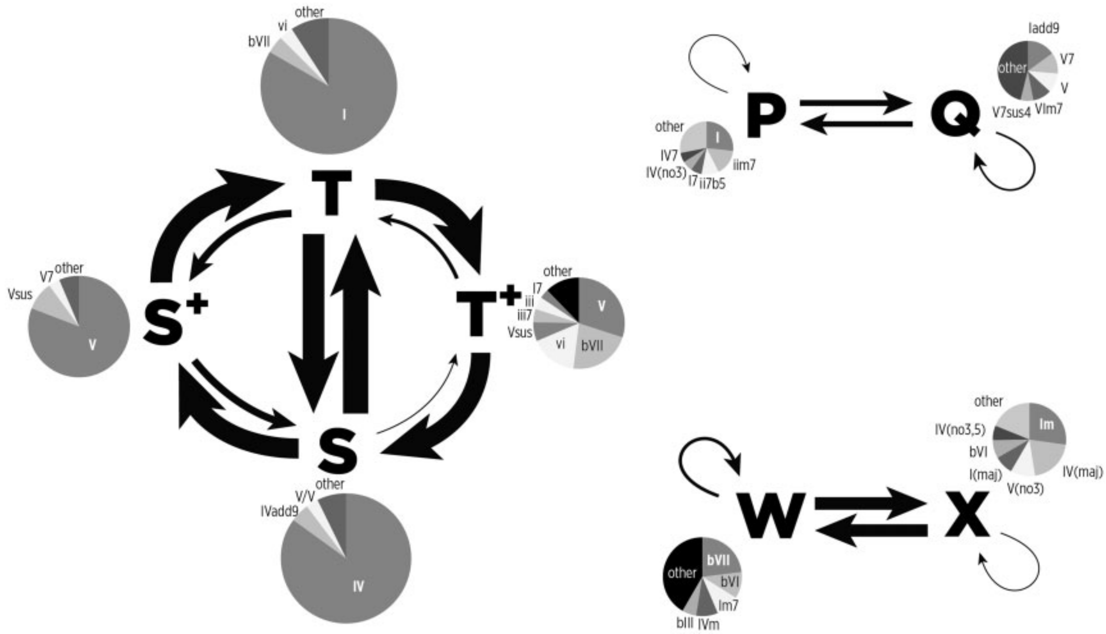
Example 10 shows a composite eight-state model of the McGill Billboard corpus. The diagram shows a main circuit of functions labeled **T, T$^+$, S,** and **S$^+$**, with two peripheral pairs, P/Q and X/W. The main circuit accounts for the substantial majority (67.2%) of chord transitions; X/W accounts for 22.7%, and P/Q accounts for only 2.3%. The remaining 7.87% of transitions are accounted for by the improbable (but possible) moves between the three circuits.[36]

The main circuit contains two primary poles, **T** and **S**, letters chosen because of the similarity between these states and the traditional tonic and subdominant functions. **T** is most frequently represented by a I chord, and **S** most often produces IV chords. These two chords are the most frequent in the corpus (23.5% and 21.7%, respectively). Furthermore, 22.3% of all transitions in the corpus are between **T** and **S**. In this sense, **S** is what we call the *antitonic* function: like the **D/T$^-$** function in the Kostka-Payne model, **S** provides the most frequent transition into and out of tonic.[37] The third most frequent chord, V, accounts for 15.5% of the corpus, and is divided between the remaining two functions in the main circuit: **T$^+$** and **S$^+$**. The defining characteristic of **T$^+$** is that it falls between **T** and **S**; the defining characteristic of **S$^+$** is that it falls between **S** and **T**. In this sense, **T+** is *pre-antitonic/post-tonic* while **S$^+$** is *post-antitonic/pretonic*. While **T$^+$** can manifest a number of chords other than V (♭VII, vi, etc.), **S$^+$** tends to produce only chords closely related to V. The distinction between **T$^+$** and **S$^+$** is thus a distinction of both syntactic position and of chord membership. The fact that there are four functions in the central circuit reflects the fact that most repeated chord progressions in this corpus are two (**T–S**), three (**T–T$^+$–S** or **T–S–S$^+$**), or four (**T–T$^+$–S–S$^+$**) chords long.

Example 11 displays three excerpts from the model's analysis of Meat Loaf's "Paradise by the Dashboard Light." The first excerpt, in D major, contains three overlapped instances of a I–IV–V–I progression, analyzed as cycles through the left side of the circuit (**T–S–S$^+$–T**). The second excerpt, also in D major, uses a different ordering of the same chords (I–V–IV–I), traversing the other half of the circuit (**T–T$^+$–S–T**): the V chord represents the **S$^+$** state in the first excerpt and the **T$^+$** state in the second. Although the contents of the two chords are identical, they have different syntactic relationships to the antitonic IV chord, with **S$^+$** as the post-antitonic function, and **T$^+$** as a pre-antitonic function. The second excerpt ends with a **T–T$^+$–S–S$^+$** progression in which the pre-antitonic **T$^+$** is instantiated with different content, a ♭VII chord; the third excerpt shows the same functional progression (now in C major) with the V chord again occupying the pre-antitonic **T$^+$** position.

Example 12 shows a chord progression that more closely resembles common-practice syntax: I–vi–II–V–I. The model interprets this progression as a clockwise traversal of the **T–T$^+$–S–S$^+$–T** circuit, with **S** now producing a II chord. (While IV is by far the most frequent **S**-functioning chord,

30 Nobile (2013).
31 Nobile (2013, 35).
32 Burgoyne (2012) and Burgoyne, Wild, and Fujinaga (2013).
33 Everett (2004).
34 This number was chosen after several pilot studies found 300 models to balance robustness of the results with the computing time required.
35 See online version of this article.
36 This contrasts with the findings of Jacoby, Tishby, and Tymoczko (2015), who find that popular music can be ideally described using the traditional three-function model using their "bottleneck" method. Their method for discovering what they call "optimal k-categorization schemes" assumes that all instances of a given chord belong to a single category. What distinguishes an HMM from a basic Markov process is precisely the possibility that two instances of the same symbol might belong to two different categories. Given that their model's underlying logic differs from ours and that the authors are committed to favoring simpler models, we believe that our findings complement theirs.
37 This reflects deClercq and Temperley's findings that IV is the most prominent nontonic chord within the harmonic language of popular music.



EXAMPLE 10. *Eight-state model of the McGill–Billboard Corpus*

the model recognizes that II occurs within the same context—between chords of the $\mathbf{T}^+$ and $\mathbf{S}^+$ functions.[38]) The progression itself bears a resemblance to the four-state model of the Kostka-Payne corpus: four functions accommodate a series of falling fifths. However, the fact that this progression expands a central **T**–**S** axis rather than a dominant–tonic pair is significant. It shows the crucial distinction between the pop/rock and the Kostka-Payne functional paradigms: while the dominant function acts *both* as pretonic and antitonic in common-practice music, these positions are distinct in the pop/rock corpus, with **S** as antitonic and either **S** or $\mathbf{S}^+$ acting as pretonic.

This paradigm then allows for cases such as Example 13. This song exhibits the same functional progression, but in a realization foreign to common-practice norms. The post-antitonic $\mathbf{S}^+$ function becomes ♭VII rather than V.[39] The four-function circuit interprets both progressions as exemplars of a more general case: the expansion of the primary **T**–**S** pair with $\mathbf{T}^+$ and $\mathbf{S}^+$ functions. Unlike the $\mathbf{T}^+$-functioning ♭VII chord within "Paradise by the Dashboard Light," the ♭VII in "I Want You to Want Me" seems to substitute for a V chord—or, more precisely, appears in a place the model recognizes that V chords often appear.[40]

The peripheral functional pairs, **Q/P** and **X/W**, arise from portions of the corpus that differ in harmonic vocabulary from the major-mode, triadic music we have examined so far. The **Q/P** pair tends to include chords involving sevenths and ninths (perhaps indicating harmonic languages more influenced by jazz). Example 14 shows the model's analysis of the Little River Band's "We Two." The alternation between I^9 and ii^7 chords shuttles back and forth between the **Q** and **P** functions. To a human trained in music theory, the scale-degree content of these chords would indicate equivalencies between I^9 and I chords and ii^7 and ii (or IV); with such equivalencies in hand, we might imagine this passage as going between the **T** and **S** functions. However, with no such equivalencies available in our strictly context-oriented approach, the model recognizes seventh and ninth chords as completely different objects from their triadic counterparts. Since fewer chords in this corpus have sevenths and ninths, and these nontriadic chords tend to progress to other nontriadic chords, the model relegates them to their own peripheral syntax, where it simply groups chords into two categories that follow one another. The tonic-antitonic polarity does not appear to hold on the periphery of the model: note, for example, that chords with roots of I, IV, and V appear in both **P** and **Q**.

The other significant lexical minority in the corpus consists of minor-mode chords, which the model relegates to the **X** and **W** functions. Example 15 shows an excerpt from "Funky Nassau," a 1971 song by The Beginning of the End that is

38 These categorizations reflect the general tendencies of groups of chords—the broader the grouping, the less the model captures the behaviors of individual chord types. The grouping of IV and II is an example of this: taken on its own, the II chord primarily moves to chords in the **S+** function, while the IV chord is more equally disposed to move toward **T** and **S+**. But regardless of these differences, the model recognizes the chords' similarity and groups them together into one function.

39 Note that ♭VII is a very improbable, but still possible, chord to be output by the $\mathbf{S}^+$ function, coming under "others" in Example 10.

40 The former functions akin to Biamonte's "double plagal" progression, and the latter as Biamonte and Doll's "dominant" ♭VII chords.

Chords	G	A	D	G	A	D
	Ain't no doubt about it, baby got to go and shout it					
R.N.s	IV	V	I	IV	V	I
Function	S	S^+	T	S	S^+	T

...

Chords	A	G	D	A	G
	Stop right there!	I gotta know right now!		Before we go any further! Do you love me?	Will you love me forever?
R.N.s	V	IV	I	V	IV
Function	T^+	S	T	T^+	S

...

Chords	D	C	G	A
	Do you need me... will you never leave me...	...will you make me so happy for the rest of my life?	...and will you make me your wife? will you take me away...	
R.N.s	I	bVII	IV	V
Function	T	T^+	S	S^+

...

Chords	C	G	F	G	C
	I started swearing to my god and on my mother's grave that I would love you to the end of time!				
R.N.s	I	V	IV	V	I
Function	T	T^+	S	S^+	T

EXAMPLE 11. *Meat Loaf, "Paradise by the Dashboard Light" (1977)*

Chords	C	Am	D	G
	Ma belle amie	You were a child of the sun	and the sky and the deep blue sea	
R.N.s	I	vi	II	V
Function	T	T^+	S	S^+

EXAMPLE 12. *Tee Set, "Ma Belle Amie" (1969)*

one of a small number of minor-key songs in the corpus. Here, **X** and **W** appear to function as tonic and antitonic categories, respectively. In general, however, these two functions do not map onto tonic/antitonic categories; rather, as is the case for **P** and **Q**, each state simply comprises chords that tend to progress to chords in the other state. For instance, major IV and minor iv tend to appear in different categories, with iv appearing opposite the tonic and IV appearing in the same category as i but opposite ♭VII.

Several interesting topics arise from these examples, particularly when comparing them to our earlier Kostka-Payne functions. Notably, because the pop-music model's parameters were learned directly from the corpus rather than adapted from preconceived three-function model, it yields a functional system with characteristics very different than that of the common-practice. The most obvious difference is the identity of the antitonic: while V functions as the most frequent non-tonic chord in the common-practice Kostka-Payne corpus, IV fills that role in the pop McGill-Billboard corpus. This is a direct consequence of the impact that chord frequency has on the functional categories the HMM learns from the corpus. The most frequent chords in the Kostka-Payne corpus are I, V, and ii, each of which becomes the nucleus of a function. In the McGill-Billboard corpus, the most frequent chords are I, IV, and V. In both cases, the second most frequent chord takes on the antitonic role, while the behavior of the remaining chords defines the dynamics surrounding the main tonic/antitonic polarity, exerting influence relative to their frequency. In

Chords	A	F#m	B	G	A
	I'd love you to love me	I'd shine up my old brown shoes		I'd put on a brand new shirt	
R.N.s	I	vi	II	^{b}VII	I
Function	T	T^{+}	S	S^{+}	T

EXAMPLE 13. *Cheap Trick, "I Want You to Want Me" (1975)*

Chords	A^{9}	Bm^{7}	A^{9}	Bm^{7}	A^{9}	Bm^{7}	A^{9}
	All alone,	on my own.		Since I walked out on you,	I walked out on me.	Now it's gone...	
R.N.s	I^{9}	ii^{7}	I^{9}	ii^{7}	I^{9}	ii^{7}	I^{9}
Function	Q	P	Q	P	Q	P	Q

EXAMPLE 14. *Little River Band, "We Two" (1983)*

Chords	Cm^{7}	F	Cm^{7}	G	Cm^{7}	F	Cm^{7}	G
	Nassau's gone funky,			Nassau's gone soul		*instrumental*		
R.N.s	i^{7}	IV	i^{7}	V	i^{7}	IV	i^{7}	V
Function	X	W	X	W	X	W	X	W

EXAMPLE 15. *The Beginning of the End, "Funky Nassau" (1971)*

the Kostka-Payne corpus, ii's frequent preparation of the dominant creates the predominant category, while in the McGill-Billboard corpus, V's role both before and after IV produce the $\mathbf{T^{+}}$ and $\mathbf{S^{+}}$ functions.

This distribution of V over two functional categories brings us to the second major difference between the two corpora: where common-practice syntax contains a "dominant" category that acts as both antitonic and pretonic, these two roles are differentiated in pop syntax, where the pretonic position can be occupied either by an **S** chord, with antitonic and pretonic functions coinciding, or by an $\mathbf{S^{+}}$ chord that acts as pretonic and usually follows an antitonic **S**.

Finally, our modeling method uses only contextual information with no knowledge of scale-degree overlap or voice-leading similarity. The fact that purely contextual data can produce a workable functional model is notable in and of itself. It is not at all obvious that a content-blind method would be able to reproduce workable categories at all. For instance, it was noted that the method models the Kostka-Payne corpus with functional categories very similar to how the authors themselves describe tonal dynamics; similarly, the pop/rock results reproduce observations made by several theorists of pop-music function.

This investigation did, however, show the difficulties that arise when analyzing a corpus containing multiple styles and practices, resulting in suboptimal peripheral functions. These cases reveal a significant weakness of the HMM approach: its underlying assumption is that a single set of probabilistic rules governs the entire training set. In this case, different segments of the corpus seem to be governed by different syntactic principles, and the resulting model is tailored only to that segment with the greatest representation. Minority segments are relegated to the periphery. We hypothesize that training separate HMMs on individual subdivisions would yield more robust syntactic models for minor-key music, jazz-influenced music, and other distinct styles. In what follows, we examine these issues in a completely uniform corpus, that of the Bach Chorales.

MODELING THE BACH CHORALE CORPUS: THIRTEEN FUNCTIONS, AND SYNTACTIC DISSONANCE

Consider Examples 16(a) and (b), two excerpts from BWV 146 and BWV 402, transposed to C major for ease of comparison. The first excerpt will give a listener familiar with Bach's chorale style momentary pause: something about the initial chord progression seems unusual. In fact, in our corpus (described below), only 21 of Bach's 2,130 V^{7} chords are immediately preceded by vi chords. In contrast, V^{7} is immediately preceded by IV 177 times: Example 16(b) therefore seems somewhat more idiomatic. Similarly, the tonic expansion in Example 16(c) seems extremely idiomatic, with the chord marked with a X, {$\hat{1}$, $\hat{2}$, $\hat{4}$, $\hat{5}$}, prolonging tonic. Indeed, this sonority occurs between two I chords 206 out of the 392 times this sonority appears. In contrast, {$\hat{1}$, $\hat{2}$, $\hat{4}$, $\hat{5}$} provides a contrapuntal prolongation of a ii^7–V^7 progression only five times, one instance of which (from BWV 244) is shown in Example 16(d).

Notably, these constraints in Bach's choices are not entirely accounted for by the rules of counterpoint. Both IV and vi lead easily to V^7, and the common tones within {$\hat{1}$, $\hat{2}$, $\hat{4}$, $\hat{5}$} make it a natural interpolation between ii^7 and V^7. If Bach's chorales contain many such style-specific regularities, then subjecting the music's surface events—e.g., not only the underlying triads, but dissonances, applied chords, and nontriadic structures—to an HMM analysis will produce a syntactic model with functional categories comparable to those modeled above, but now tailored specifically to Bach's chorales.

MATERIALS AND METHODS OF THE BACH CHORALE CORPUS

Before implementing the HMM, we needed to determine what kind of computer-readable representations of the Bach chorales were to be processed and how to divide these representations into observations. We began with the chorale corpus distributed with the music21 software.[41] In order to remain as close as possible to the musical surface, we parsed the surface not into Roman numerals or chords but *salami slices:* each time the pitch-class content of the chorale changes by adding or subtracting one or more notes, a new salami slice is identified, consisting of a snapshot of all sounding pitch-classes.[42] Salami slices are represented as unordered sets of scale degrees relative to the local key, which was determined using a preexisting key-finding process.[43] The result was 35,139 salami slices, of which 1,079 were excluded for tonal ambiguity.

To assign a key to these files (and to identify modulations), we subjected the corpus to a preprocessing step. The process, described in White (2013), used a moving window of eight chords; the pitch-class content of each window was analyzed by a key-profile analysis implemented in music21. The two most likely keys for each window were determined together with the algorithm's confidence in both keys (on a scale of -1 to 1) and, if the confidence value for those two keys differed by less than a threshold amount (0.1), the window was determined to have ambiguous tonality. The key of each chord was then determined by comparing the eight windows containing that chord (excluding ambiguous windows). The key of the window having the highest confidence value was taken as the key of the chord. Stretches of chords in a single key were used as the observations for the HMMs.

Unlike the constrained vocabularies of the earlier examples (the Kostka-Payne corpus uses 12 chords, the McGill-Billboard uses 68), this method produced a vocabulary of 329 distinct salami-slice types. Relative to the corpus size, this vocabulary was too large for the Baum-Welch procedure to produce sufficiently consistent results. Initial experimentation with the corpus suggested three simplifying steps to reduce the number of distinct slice types within the observation sequences. First, we considered only major-mode stretches of music (totaling 22,569 slices), to avoid the problem of "lexical minorities" discussed above in connection with the McGill-Billboard corpus.[44] Second, we ignored dyads and singletons that immediately preceded or followed proper supersets of themselves: a $\hat{3}$–$\hat{5}$ dyad following a complete I chord would be ignored. This step removed 5.6% of the salami slices. Finally, low-probability slices were removed from the observation sequence, since removing these chords dramatically improved the HMMs' performance in pilot studies. Slices with a frequency below the 40th percentile were removed, which amounted to only 2.3% of the total corpus.[45]

The HMM procedure was otherwise identical to that used with the McGill-Billboard corpus.

RESULTS

The values of *k* yielding the most consistent analyses of the Bach chorales were 3 and 13. Example 17 schematizes the three-function solution (specific lexical and transition tables can be found in Appendix B), which closely resembles the three Riemannian functions. I, V, and IV are the primary chords for the three hidden states, and, just as we might expect, vi and iii stand in for tonic in some situations, while vii^{o} sometimes acts as dominant, and several chords—ii, ii^7, or vi—act as predominant. As in traditional tonal theory, the preferred order of functions is **T–P–D–T**. Example 18 shows how the model reads a I–ii^7–V$^{4-3}$–I cadence as just such a functional progression.

While the model conforms to many of our intuitions about harmonic function, it has some counterintuitive features that can be attributed to the strictly context-based analysis of salami slices. In Example 18, for instance, the **D** function does not arrive until the last beat of the first full measure. This is because the model recognizes that V chords with a suspended fourth always precede V chords and are, therefore, categorized in the **P** function. Example 19(a) and

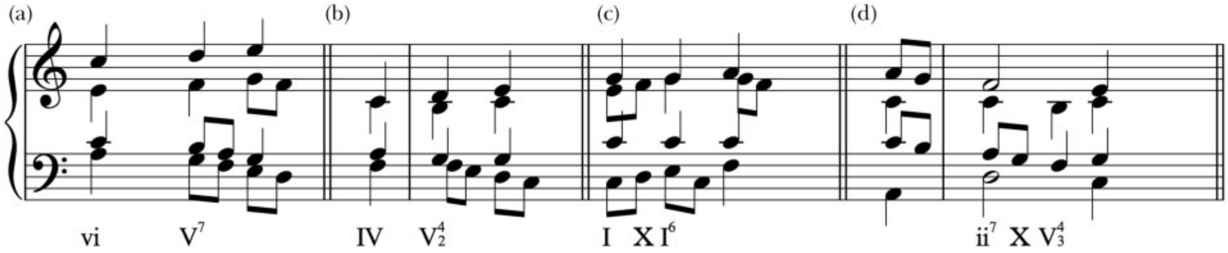
41 Cuthbert and Ariza (2011).

42 Quinn (2010).

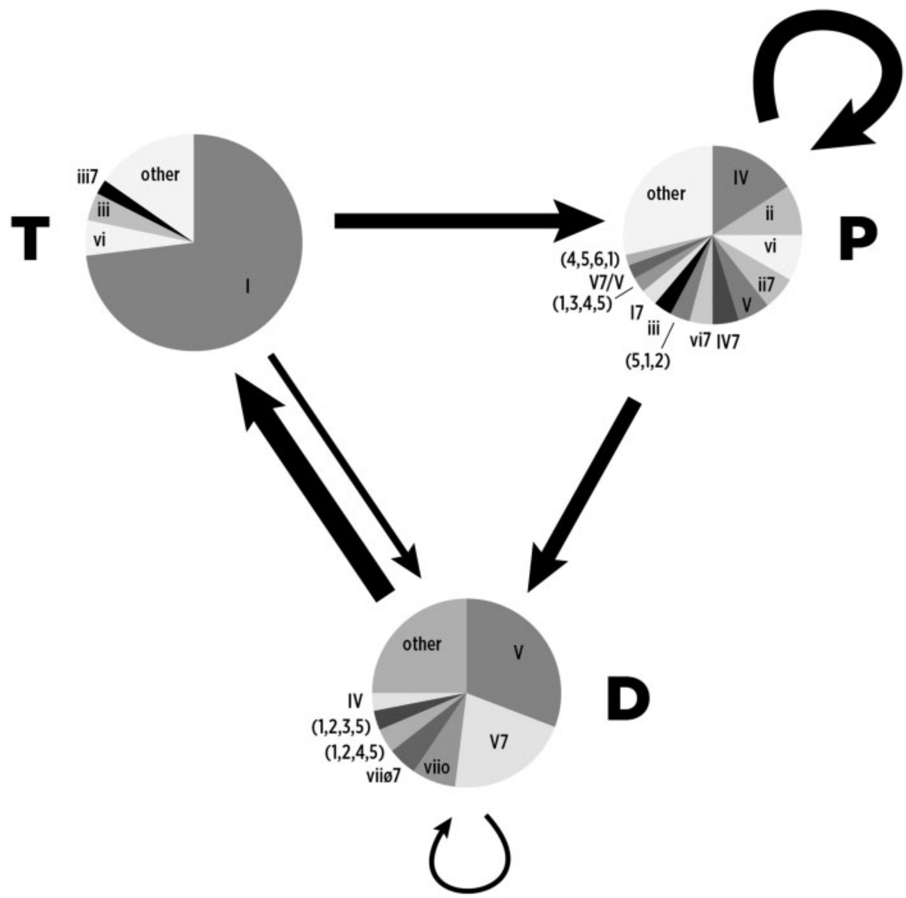
43 White and Quinn (2016). Note that this representation does not distinguish between chords with different voicings, doublings, or inversions: only scale-degree content matters. For example, the first two beats of Example 16(a) contain three salami slices: assuming the key of C major, the first quarter note's slice would be the set <0, 4, 9>, then the next two eighth-note pulses would be the sets <2, 5, 7, 11> and <2, 5, 9>, respectively.

44 This excluded the 33.7% of the corpus in the minor mode and removed 29.4% of the chord types.

45 This procedure is frequently used in data sets with "long tails." See Quinn and Mavromatis (2011) for such a discussion.



EXAMPLE 16. *(a) BWV 146b, (b) BWV 402, (c) BWV 245, (d) BWV 244*



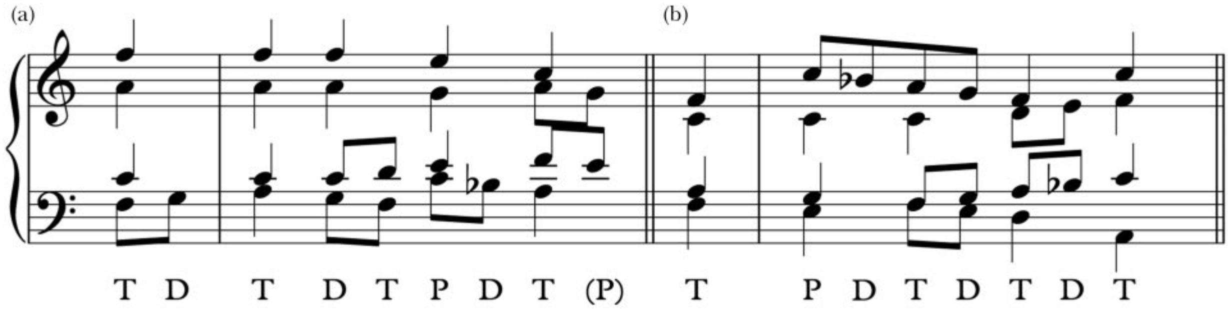
EXAMPLE 17. *The three-function model*

(b) illustrate further peculiarities of the model. In Example 19(a), apparent fourth-inversion I^9 chords—verticalities that pass between two tonic triads—function as **D**, since these sorts of chords always progress to **T** chords. Notice that even though the second of these I^9 chords does not proceed to I, it does go to vi, a chord that may function as **T**. Similarly, the penultimate verticality of Example 19(b), $\{\hat{1}, \hat{4}, \hat{6}, \hat{7}\}$, always proceeds to I or vi chords and is, therefore, analyzed as **D**. Both excerpts also include "predominant" V chords: the model recognizes that V often precedes chords that themselves precede **T** chords. In both examples, the V chord is followed by the addition of $\hat{4}$ to the texture, creating the vii° and V^7 slices, respectively.

The three-function solution, then, produces a model with traditional tonal relationships along with some nontraditional instantiations of those relationships. The asymmetry of the transitions between functions is familiar: the **P–D** transition is unidirectional, while **T–P** and **T–D** transitions are basically bidirectional, with a disposition for **T–P–D–T** motion.[46] However, the model sometimes applies these functions and

46 This results in a very messy predominant function. While 70% of tonic chords are I triads and more than 35% of dominant chords are V triads, the most frequent predominant chord, IV, only instantiates that function around 16% of the time.

EXAMPLE 18. *BWV 10.7, mm. 12–14*



EXAMPLE 19. *(a) BWV 127.5, m. 1, (b) BWV 10.7, m. 11*

transitions to slices that we would not ordinarily consider as function-bearing chords.

A distinct advantage of the thirteen-function solution (see Ex. 20) is that it adds new functional categories that accommodate such slices. This model retains the basic tonic-centric circular flow we have seen in many of our HMM solutions, but adds parallel pathways and detours. Example 21 illustrates some idiomatic progressions of these pathways. The top of Example 20 shows **T** (*tonic*), a function which contains mostly I triads along with several vi and iii chords. $\mathbf{T^x}$ (*tonic expansion*) is composed of slices resulting from passing and neighbor motion between two **T** slices, including the nontriadic $\{\hat{1}, \hat{2}, \hat{4}, \hat{5}\}$ of Examples 16(c) and 21(a). Example 21(b) shows the primary pathway of "strong" functions around the outside of the diagram, so named because of their phrase-ending cadential function. Here, **T** first moves to **P** (*strong predominant*), to **D**—the *strong dominant*—and then to $\mathbf{D^+}$, the *late dominants*, comprised mostly of V^7). The cadential progression I–ii–V^{8-7} would traverse this outer pathway, as would the V-chord suspension of Example 21(b).

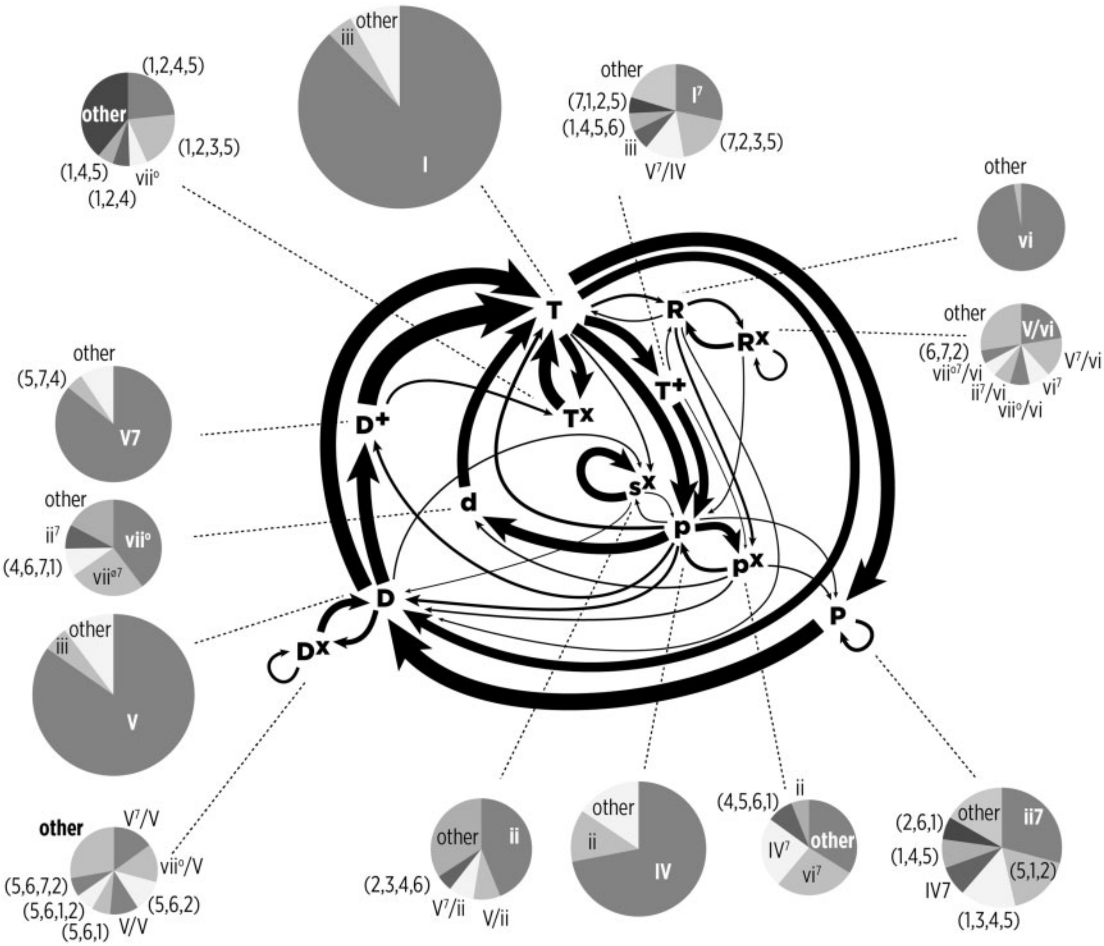
The inner cycles show progressions that tend to precede cadential progressions, or the "weak" functions. Example 21(c) shows such a progression. First, the I^4_2 chord functions as a *late tonic* $\mathbf{T^+}$, a passing function that progresses from **T** to **p**, the middle-of-phrase *weak predominant*. The $\mathbf{p^x}$ function (*predominant expansion*) then prolongs the **p** function with its passing chord. Finally, the model involves three tonicization detours. Example 22 shows how V, vi, and ii can receive their own tonicization functions, and Example 23 summarizes the thirteen functions.[47]

Example 24 shows the models in analytical action. The three-state analysis assigns hidden states in a way that conforms to our intuitions, moving through two complete **P–D–T** cycles. Functions are sometimes instantiated by more than one chord, and the tonic of the first cycle is expanded by a passing dominant. The thirteen-state analysis distinguishes between the nontonic functions of the first measure and those of the second. The initial measure uses the weak predominant (**p**) and dominant (**d**) functions, and both **p** and **T** are prolonged by their corresponding expanding functions. The cadential part of the phrase uses the strong predominant (**P**) and dominant (**D**) functions. The phrase, on the whole, moves first through the inner loop of Example 20, and then through the outer loop.

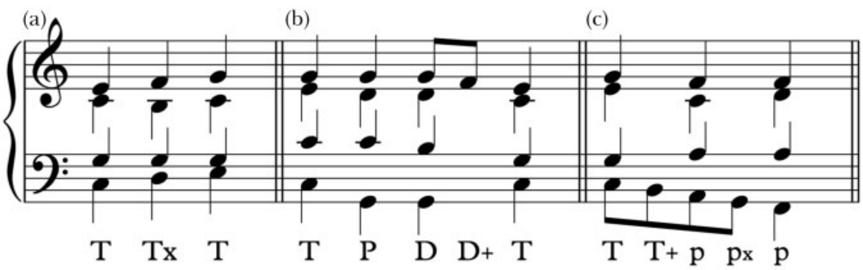
Example 25 investigates several other functions within the thirteen-state model. The phrase as a whole begins in E major and leads to what turns out to be a fleeting tonicization of the relative minor. Our model's thirteen-state analysis shows three functions (**T**, **p**, and **R**), each prolonged by their expanding functions. The tonicization of the submediant is treated as a within-key phenomenon.

Example 26 highlights the remaining functions available to the thirteen-state system. The first measure involves a tonic

47 There are two further important characteristics concerning the x/ii function. First, in the thirteen-state model, this function involves both ii and its applied chords, while a fourteen-state model divides these two categories (just like R and R^x in the thirteen-state model). This functional division is the primary difference between thirteen- and fourteen-state models. Second, it is notable that the models distinguish between ii as a tonal area and ii as a predominant chord.



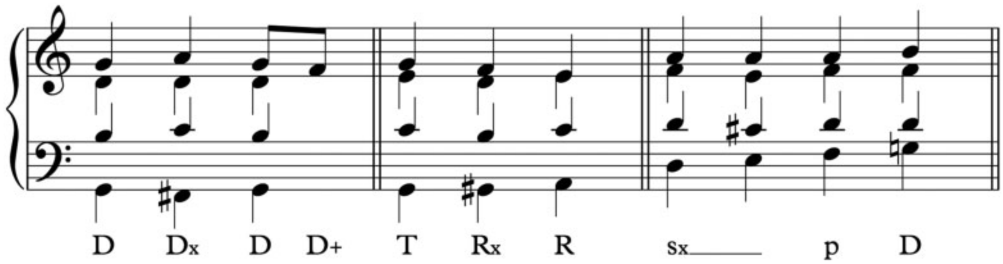
EXAMPLE 20. *The thirteen-state model*



EXAMPLE 21. *Some representative instantiations of the thirteen functions: (a), (b), and (c)*

expansion using first a late tonic followed by a **p**-functioning iv^7 chord.[48] The second measure involves a dominant

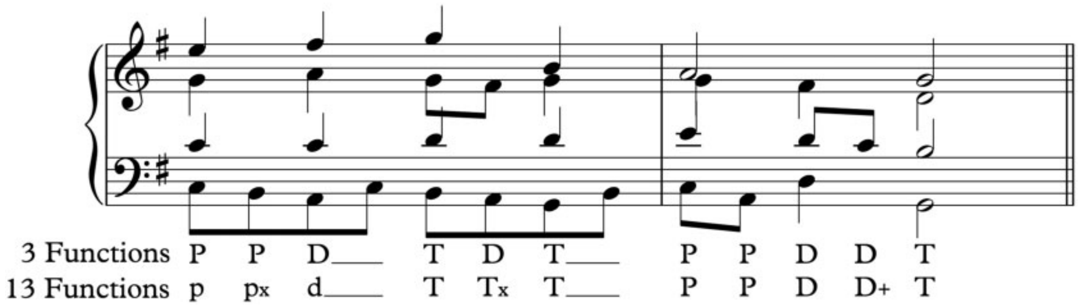
48 Note that the following passing I chord is labeled as a **T**: this is because the model labels all I chords as **T** due to the overwhelming frequency of I chords occurring in the early tonic context. Another notable issue of the model struggling to analyze a I chord involves cadential $^6_4$ chords. On one hand, the fact that the model does not recognize bass notes might account for the model's apparent difficulty with I chords preceding cadences. But, if this progression happened often enough, we would imagine that the HMM would categorize those I chords that follow ii and IV chords and precede V7 chords as **D** or **P** functions. However, Bach simply does not use enough cadential $^6_4$s for the model to incorporate this behavior. Only 30 of Bach's 455 I–V progressions are in the proper inversion to be considered cadential candidates.



EXAMPLE 22. *Some further representative instantiations of the thirteen functions: (a) and (b)*

function	*characteristic chords*	*behavior*	*corpus probability*
T	I, iii, vi	Comes from **D**-functions, goes many places. Expanded by $\mathbf{T}^{x}$.	19.0%
D	V, iii	Comes from **P**-functions, **T**, and **R**, goes to $\mathbf{D}^{+}$ and **T**. Expanded by $\mathbf{D}^{x}$.	18.4%
$\mathbf{D}^{+}$	V^{7}, all its trichord subsets	Comes from **p** and **D**. Goes to **T**.	13.2%
P	ii^{7}, $\{\hat{1}, \hat{2}, \hat{5}\}$, $\text{I}^{\text{add 4th}}$	Comes from other **P**-functions and **T**. Goes to **D**. Expanded by itself.	9.0%
s^{x}	ii, $\text{V}^{(7)}$/ii	Comes from **D** and **T**. Goes to **p**. Expanded by itself.	8.4%
$\mathbf{D}^{x}$	V^{7}/V, all its trichord subsets	Comes from and goes to **D**. Expanded by itself.	7.8%
p	IV, ii	Comes from **T**-functions and s^{x}. Goes to **D**-functions. Expanded by $\mathbf{P}^{x}$.	6.0%
$\mathbf{R}^{x}$	V^{7}/vi, all its trichord subsets	Comes from and goes to **R**. Expanded by itself.	5.9%
$\mathbf{T}^{x}$	$\{\hat{1}, \hat{2}, \hat{4}, \hat{5}\}$, $\{\hat{1}, \hat{2}, \hat{3}, \hat{5}\}$, $\{\hat{1}, \hat{2}, \hat{4}\}$, vii	Comes from and goes to **T**.	4.9%
d	vii^{o}, $\text{vii}^{\text{ø}7}$	Comes from **p** and $\mathbf{p}^{x}$, goes to **T**.	3.6%
$\mathbf{p}^{x}$	vi^{7}, IV^{7}	Comes from and goes to **p**, but also can come from $\mathbf{T}^{+}$ and **R** and go to **P** or **D**-functions.	2.0%
R	vi	Comes from and goes to **T** and **R**, but can also go to $\mathbf{P}^{x}$ and **D**.	1.3%
$\mathbf{T}^{+}$	I^{7}, $\{\hat{2}, \hat{3}, \hat{5}, \hat{7}\}$, $\text{I}^{\flat 7}$	Comes from **T**, and goes to **p** or $\mathbf{p}^{x}$.	0.7%

EXAMPLE 23. *A summary of the thirteen functions*



EXAMPLE 24. *BWV 115.6, mm. 3–4*

EXAMPLE 25. *BWV 124.6, mm. 8–9*

expansion and late dominant function, and the final eighth note of that measure shows the "early tonic" capacity of the $\mathbf{T}^{\mathbf{x}}$ function, intervening between the cadential V^7 ($\mathbf{D}^+$) and the cadential tonic (**T**). The second half of the third measure briefly tonicizes the minor supertonic, something the thirteen-state model labels as a patch of $\mathbf{s}^{\mathbf{x}}$ states.[49] This model treats tonicizations of ii differently from tonicizations of V and vi. Each of the latter two *Stufen* has separate states for the tonicizer and the tonicized, but tonicizations of ii are rare enough that the model only learns a single state for all chords involved in tonicizations.

COMPARING AND CONTRASTING THE MODELS: THREE GENERALIZATIONS

The differences between the models makes it difficult to generalize about context-oriented harmonic function, and the fact that we present only three corpora makes generalizations about musical syntax even more suspect. Although our discussion so far has primarily emphasized the differences between the contextual regularities of each corpus, our models have some commonalities that are shaped both by the properties of the corpora and by the HMM method itself. The expectation-maximization procedure for training the HMMs produces a model that assigns a high probability to observation sequences in the training set, and this preference favors models that make strong predictions about the corpus's most frequent chords. High lexical probabilities associated with common chords and high transition probabilities associated with common hidden-state progressions will result in high probability estimates for observation sequences. In other words, because a hidden state that is equally likely to produce any of a handful of observable chords will result in low probability estimates for any of those chords, the Baum-Welch algorithm prefers high lexical and transition probabilities associated with the most frequent hidden states. This corresponds to a minimization of lexical diversity in the most common hidden states: note that the larger pie charts in the examples tend to have fewer, larger slices.

However, several characteristics arise neither because of the properties of HMMs nor the Baum-Welch algorithm, but because of the properties of the corpus. These recurrences are notable not only because they arise within all the corpora investigated, but because there is no reason to expect our methods to produce models with these characteristics.

First, each model's top two most frequently occurring chords create two functions that strongly associate with one another in some way (e.g., the I and V chords in the Kostka-Payne and Bach corpora, and the I and IV chords in the

49 Recall that $\mathbf{s}^{\mathbf{x}}$ also involves ii itself. In the fourteen-state model, $\mathbf{s}^{\mathbf{x}}$ is divided into two states, one of which is solely ii, while the other comprises applied chords that move to ii.

EXAMPLE 26. *BWV 157.5, mm. 5–8*

McGill Billboard corpus). In each case, the most frequent chord, I, creates a recognizable tonic function and, as such, creates an important pillar in the functional system. The second most frequent chord, then, provides a secondary pillar, and its relationship to tonic creates one of the most defining transitions of the model. We will refer to this general property as the *tonic/antitonic dichotomy*: each corpus has two hidden states built around two most frequent chords in the corpus.

Second, each corpus has a unidirectional relationship between a state or states that move into tonic but to which tonic progresses less frequently. We could imagine this as the "cadential progression"; however, not wanting to import the meanings inherent in that term, we will adopt the term *pretonic/tonic relationship*. In the Kostka-Payne and Bach corpora, this dynamic involves the dominant states progressing to the tonic states, and in the popular music corpus the $\mathbf{S}^+$–**T** progression instantiates this relationship. Notably, while these dynamics overlap with the tonic/antitonic dichotomy for the Bach and Kostka-Payne corpora (the dominant-tonic "cadence"), the two relationships do not overlap in the pop-music corpus: the unidirectionality of the $\mathbf{S}^+$–**T** "cadence" is not the same as the **S**–**T** bidirectional tonic/not-tonic dichotomy.

Finally, each model contains less-probable pathways that, rather than constituting their own systems, expand the primary tonic–antitonic or tonic–pretonic pair. These pathways comprise the *prolongational networks* that either act as precursors or successors to the model's most frequent pathways. For instance, adding $\mathbf{S}^+$ between **S** and **T** in the McGill-Billboard model prolongs the tonic–antitonic dichotomy similar to the way that the **P** and $\mathbf{P}^-$ prolong the **T**–**D**/$\mathbf{T}^-$transitions of the Kostka-Payne model. On a larger scale, we could even imagine the "weak" inner pathways of the thirteen-state Bach model prolonging the cadential "strong" progressions of the outer pathways that provide the model's antitonic/pretonic function.

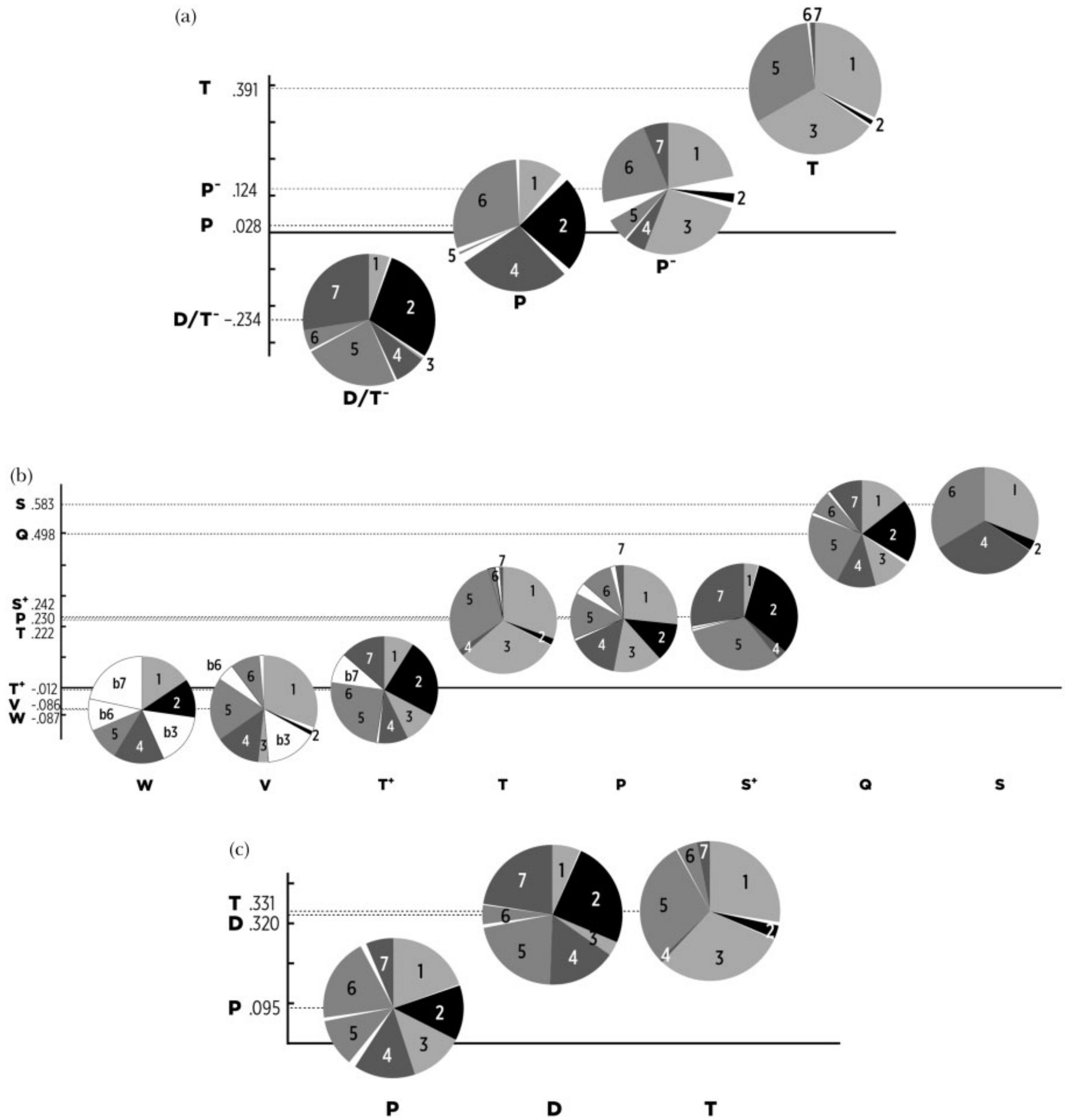
COMMON TONES, FREQUENCY, AND WHAT IT MEANS TO BE A FUNCTION

Example 27 shows several pieces of information concerning the scale-degree content of each corpus's hidden states. The pie slices show the proportional frequency with which each scale degree occurs in each function: pies with fewer, larger slices are dominated by fewer scale degrees. The pies are ordered from left to right by their ascending *common-tone* scores, indicated on the *y* axis. The score is a measure of the degree to which chords in the function share scale degrees. It is not a raw count of common tones, but a scaled and normalized measure: a score of zero corresponds to the average number of common tones between any two chords randomly chosen from the corpus. A positive score for a function means that chords within that function have a greater than average number of common tones; a negative score means a lower than average number of common tones. The score is scaled in terms of standard deviations of the distribution over the corpus as a whole (what statisticians call a *z-score*).[50]

Nearly every function in each corpus has a positive common-tone score, indicating that a small number of scale degrees dominates each function. In the Kostka-Payne corpus, for instance, the **T** hidden state is dominated by $\hat{1}$, $\hat{3}$, and $\hat{5}$, while the **P** hidden state uses primarily $\hat{1}$, $\hat{2}$, $\hat{4}$, and $\hat{6}$. At first glance this is not a surprising finding; however, recall that our models are strictly context-based and thus entirely blind to scale-degree content. In fact, a closer look at the distributions seems to suggest that certain scale-degree combinations seem to connote certain functions. In the Kostka-Payne corpus, for instance, $\hat{4}$ prominently occurs in two functions; however, when it occurs alongside $\hat{6}$ it would likely appear in the **P** function, while if it appeared with $\hat{7}$, that sonority would likely be classified in the **D** function.[51] Despite this limitation, our models reveal a deep connection in our corpora between a chord's scale-degree content and its contextual tendencies. Based on these results, the basic principle of harmonic function that identifies function with scale-degree content seems to hold in most cases.

50 The process did not compare identical chords to one another. Our test is designed to ask the question, "How much overlap exists between nonidentical chords within a function?", removing any effect that a single chord's overwhelming frequency within a particular function might have. We have represented this property in terms of standard deviations (a z-score) in order to use the same scale between corpora.

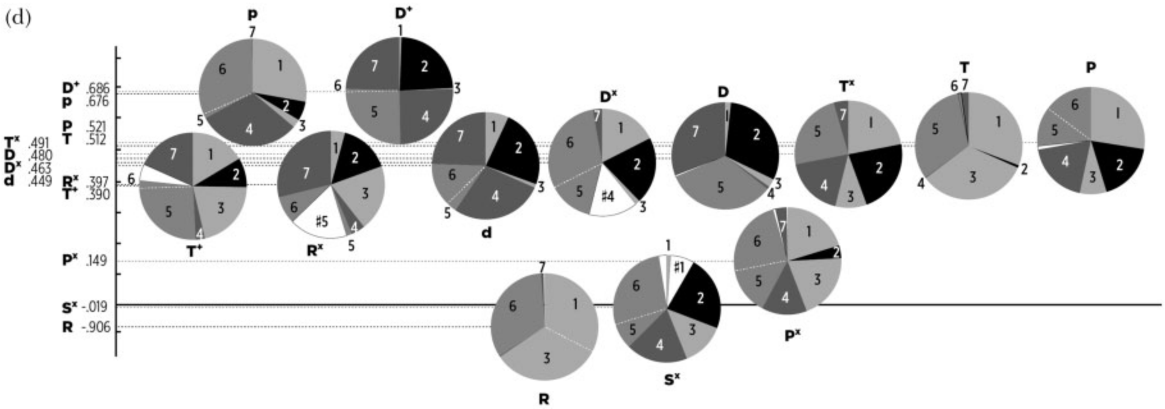
51 This topic is taken up in more depth in Quinn (2017).



EXAMPLE 27. *Common tones in each functional model: (a) The Kostka-Payne Corpus; (b) The McGill Billboard Corpus; (c) Bach Chorale Corpus, 3 states; (d) Bach Chorale Corpus, thirteen states*

But content and context do not always coexist peacefully. The **D/T⁻** category in the Kostka-Payne corpus has a lower-than-average common-tone score: chords that progress to **T** can have various scale-degree structures. Nowhere is this more evident than the "dominant" IV of the Kostka-Payne corpus: our models' analysis in this case suggests that a completely contextual definition of function might conflate traditional subdominant and dominant functions. The identical logic applies to the $\mathbf{T}^{+}$ function in the popular music corpus, and the $\mathbf{s}^{\mathbf{x}}$ and **R** functions in Example 20: in all these cases, the diversity of chords within these categories creates a disjunction between a content-oriented theory and our context-oriented model. These cases are, however, in the minority: in fact, context- and content-oriented functional theories seem to overlap more than we might at first expect.

But what of the hierarchies and qualities we usually associate with functions? For instance, we usually couple tonic with the quality of resolution and accept it as the deepest structural harmony. Our results support previous research that connects these sorts of qualities and hierarchies with the frequency with which we hear stimuli. Over the past several decades, the work of many researchers in the field of music cognition, most notably that of Krumhansl, Huron, and Aarden,[52] has made a compelling case that our perceived hierarchy of scale degrees is



EXAMPLE 27. *[Continued]*

due to the frequency with which we hear them in music, especially at the ends of phrases. Their research has shown that the most frequent scale degree, tonic, is imparted with a feeling of resolution and hierarchical superservience; in contrast, chromatic scale degrees occur very infrequently and feel hierarchically subservient, unsettled and unresolved.[53] It follows that the same might be true for harmonic functions. Our model's tonic categories may be more fundamental or structural simply because they are more frequent. Antitonic chords may be viewed as hierarchically secondary inasmuch as they participate in the second-most-frequent function. Similarly, the pretonic functions are only hierarchically subservient or points of highest tension because they precede the most-frequent function. In sum, relative frequency produces a sense of relative stability.

CULTURAL AND PEDAGOGICAL IMPLICATIONS

Importantly, we are advocating for a theoretical framework rather than any particular instantiation of that theory. We believe that contextual modeling can provide insights into the concept of harmonic function but remain open to revisions, caveats, and addenda to our specific models. Our findings therefore apply more to how we approach teaching or discussing harmonic function rather than to precisely what to teach or discuss. For instance, we readily acknowledge that any classroom-oriented theory of harmony should involve bass lines and melodic progressions, both absent in the current models. However, we will close by offering two recommendations concerning the general discourse surrounding harmonic function.

First, our models and analyses have shown degrees of subtlety, stylistic regularities, and syntactic categories that are not captured by traditional three-function harmonic theory. This can stem from the decoupling of proteinic and antitank roles (as in the popular music repertoire). It can also result in the division of predominant into more than one syntactic category (as in the **P**$^{-}$ chords in the Kostka-Payne corpus), the articulation of unique passing and neighboring functions (as in the Bach corpus), or even some other contextual regularity not addressed in the current work.

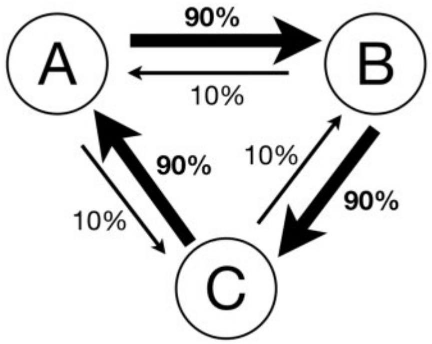
Second, we suggest that discussions of harmonic function be moored to particular repertoires and be sensitive to differences between corpora. As theorists and teachers, we should emphasize the syntactic norms of a corpus rather than universal rules. Instead of a tonic–subdominant–dominant paradigm that reaches from the Baroque to the Beatles, we might emphasize the constrained categories of chords that Bach deploys at phrase endings, versus the recurrent riffs in rock music. Assuming the three Riemannian functions as harmonic universals is at best an oversimplification and at worst culturally hegemonic. Importing a model associated with the German-language common practice inspired by Hegelian dualism onto other culturally specific repertoires problematically asserts the power of one culture over another. By deriving our models from preexisting and wide-ranging corpora, our methods provide a way to sidestep the difficult implications of using a "common practice" as the main driver of theory and as a yardstick for tonal norms.

52 Krumhansl (1988, 1990); Huron (2006); and Aarden (2003).
53 See Huron (2006) for a complete list of scale-degree qualities and the statistical rationale for each.

APPENDIX A

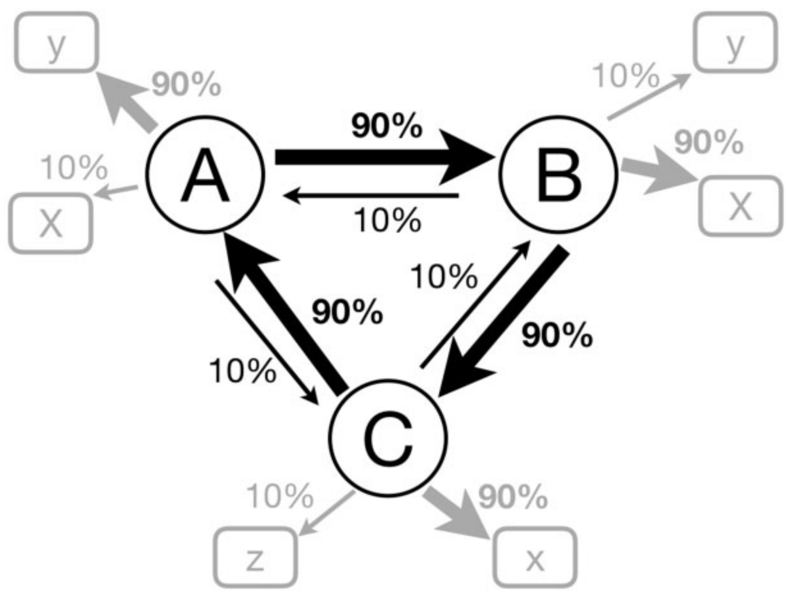
In this appendix, we present the article's formal models and algorithms in more details, including *Hidden Markov Models*, *expectation-maximization*, and *silhouette width*. Mathematical formalizations of these processes can be found in Jurafsky and Martin (2008).

Example A1 shows a simple Markov model that uses three states: A, B, and C. Each letter progresses to the next in the alphabet with a 90% probability, and backward only 10% of the time. We can use the model to assess the probability of some sequence: the progression <A, B, C> would be 72.3% probable, (.9 x .9 x .9), whereas <A, B, A> would be 8.1% likely (.9 x .9 x .1). Markov models are useful at assessing the probability of series of simple observations. If we, say, knew how likely individual chords were to progress to one another, we could assign a probability to sequences of chords, and therefore know which sequences were more probable than others given the model.

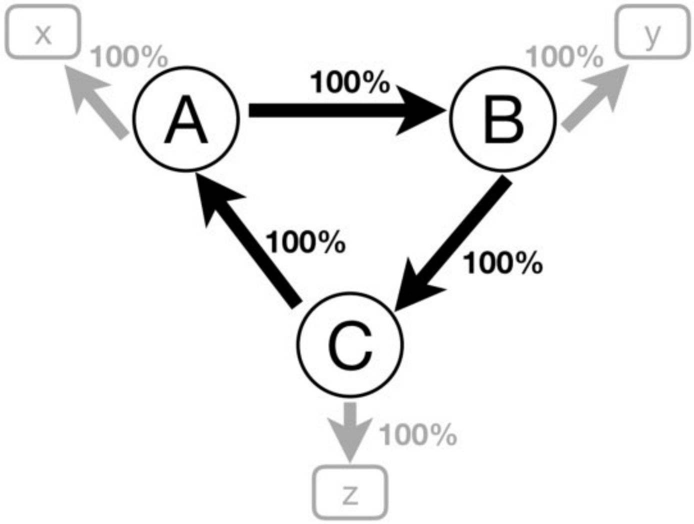


EXAMPLE A1. *A Markov Model*

Example A2 shows a Hidden Markov Model. Added to the Markov model are several *observations*—x, y, and z—and A, B, and C are now the *hidden states.* The transitions follow the same logic as before: the probability of the hidden-state sequence <A, B, A> would still be 8.1%. However, the observations add another layer. As in this article's functional analyses, such a two-layered model is used to make judgments about something you can't currently observe using something you can. For instance, economists might use observable weather temperatures to predict demand for air conditioners; a spell-checking model might observe a typed misspelling and then replace it with the intended autocorrected word. In these models, the hidden state depends on the *lexical probabilities* connecting the hidden state and the observation (i.e., how likely is a hot day to inspire air conditioner sales, or is it more probable that I meant "soup" or "soap" when I typed "can of sop"?). In HMMs, the hidden states also depend on the preceding hidden states, or the *transition probabilities*. For instance, today's air conditioner sales are influenced by yesterday's, and the best autocorrected word is dependent on the previous words. In Example A2, if we observed the string <x, y, x>, there are six options for hidden-state sequences: <A, B, C>, <A, B, A>, <B, A, B> <C, B, C>, <C, A, C>, and <C, A, B>. If we assess the probability of each, given the model's lexical and transition probabilities, we would realize that <C, A, B> provides the highest combined probabilities. It is the only option that traverses only 90% pathways. The maximum probability of the sequence <x, y, x> given the model would be the product of transition probabilities of the sequence <C, A, B> (or, .9 x .9 x .9) and the probabilities of x, y, and x being produced by those respective hidden states (again, .9 x .9 x



EXAMPLE A2. *A Hidden Markov Model*



EXAMPLE A3. *A Hypothetical Improvement to Example A2*

.9). The reader is encouraged to experiment with the model, proposing observations and adducing the hidden states.

HMMs can also derive their parameters from a body of data. A model can compile lexical and transition probabilities by tallying how frequently observations accompany hidden states and how often the hidden states follow one another. Given a body of data connecting air conditioner sales with temperature or typos with intended spellings, a programmer could model those connections using an HMM.

However, in the current article, we begin only with a series of observations—chords—and attempt to find a model to assess the unobserved hidden states—functions—not even knowing how many hidden states underlie the observations. To do this, we use an iterative process of trial and error to determine which models best explain the observation series, called the *training corpus*. One could imagine, for instance, that if we tried to explain an observation sequence that only ever moved from x to y to z back to x, the model in Example A2 would perform poorly. The (admittedly trivial) model of Example A3 would return higher probabilities, namely always 100%. This process of tweaking the transition and lexical probabilities within a model to increase the fit between the model and the observations is called *expectation maximization*. In this paper, we use the Baum-Welch expectation maximization algorithm, which cycles through possible lexical and transition settings, trying to improve the probability of the analysis given the model for each iteration. The algorithm returns a final model either after a certain number of cycles, or after it starts returning diminishing improvements (this being a parameter that can be set by the analyst).

After training, the HMM can then be used to analyze a *test corpus*. During the testing phase, an HMM assigns the most probable Hidden States to the test corpus' observation string. In order to claim that a model is generalizable and not over fitted to the peculiarities of a particular observation string, the training and test corpus should not be the same data set. For the current article, we use the Viterbi decoding algorithm for this task (again, formalizations appear in Jurafsky and Martin 2008).

As an example, consider the "good" two-state HMM and toy corpus in Example A4. The example includes (a) the lexical probability table, (b) the transition probability matrix, and (c) the probabilities assigned by model to the observation string. Here, the model includes a category *A* that has a 100% probability of manifesting I chords, a 50% probability to transition to itself, and a 50% probability of transitioning to state *B*. *B*, on the other hand, shows V half the time and V^7 the other half of the time; *B* transitions to *A* 100% of the time. The probability for the sequence would be the product of each lexical and each transition probability. (Note that V and V^7 are grouped into the same hidden state not because of their common tones, but because they occur in the same contexts.) Compare this model to Example A5, and the "poor" two-state HMM. The toy solution now groups V^7 with I in state A, and both states move between one another with equal probability. The transition and lexical probabilities of observed and hidden states given the model are lower (worse) than those in the Good two-state solution. Because the good solution has a higher probability than the poor solution, an expectation maximization process would discard the latter in favor of the former. In our modeling, we will prefer solutions like Example A4 over Example A5 for this reason.

While the comparison between the Good and Poor two-state models illustrates how the algorithm chooses between different possible solutions, the process also must choose between different numbers of hidden states. That is, how would

a) The HMM's Lexical Probabilities:

State		A	B
Observed Chord	I	100%	0%
	V^7	0%	50%
	V	0%	50%

b) Transition Probabilities (shown as the probability of row moving to column)

	A	B
A	50%	50%
B	0%	100%

c) The HMM's analysis of a toy corpus

Observations:	I V^7 I I V I I I V^7 I I V I I I V^7 I
Good 2-State:	A B A A B A A A B A A B A A A B A
Lexical Probs:	1 .5 1 1 .5 1 1 1 .5 1 1 .5 1 1 1 .5 1 = .0625
Transition Probs:	.5 1 .5 .5 1 .5 .5 .5 1 .5 .5 1 .5 .5 .5 1 = .00049

EXAMPLE A4. *Good two-state HMM analyzing Toy Corpus no. 1*

we know that two states optimally underpin the observation string, as opposed to three, four, or more?

Consider the two hypothetical Good four-state analyses in Example A6. Both distinguish between I, V^7, and V as different functional entities (states A, B, and C), and both identify the fourth state, D, as one which outputs I chords that occur in particular relationships to states B and C. In the first model, D occurs before B or C, and in the second it occurs after. If the lexical and transition probabilities were constant between both models and the latter only differed from the former in whether D proceeds to or from B and C, the probability of the sequence would be identical between both models. It would not be clear which model represented the better option for an expectation maximization procedure. Because there is no obvious solution, the algorithm would sometimes produce the first solution, and sometimes the second. This lack of consistency suggests that four is not an ideal number of states to produce this observation sequence. In our modeling, we will prefer solutions like Example A4 over Example A6 for this reason.

To quantify this, we use a measurement called *silhouette width*, a calculation used in cluster analyses. The silhouette width of point i can be found first by computing $a(i)$, the average distance between i and each other point in its cluster. The $a(i)$ therefore tells us how well i is matched with other points in its own cluster. We then compute $b(i)$, the average distance between i and the points in its closest neighboring cluster. The $b(i)$ therefore tells us how distinct i is from its closest cluster. The silhouette width then subtracts $a(i)$ from $b(i)$ and divides by the larger of the two. Assuming the point is situated in the best cluster, the number will always be between zero and one (if the point is in the wrong cluster the number will be below zero). The closer the value is to one, the better the fit to its cluster. For an overall measure of an entire clustering solution, the average silhouette width simply averages of the silhouettes of all the individual points.

The current article uses each model's analyses of the hidden states of the test corpus to create our cluster analyses. For each chord in the test corpus, every HMM with a particular cardinality of hidden states assesses the chord's state, and

a) The HMM's Lexical Probabilities:

State		A	B
Observed Chord	I	80%	20%
	V^7	20%	0%
	V	0%	80%

b) Transition Probabilities (shown as the probability of row moving to column)

	A	B
A	50%	50%
B	50%	50%

c) The HMM's analysis of a toy corpus

Observations:	I V^7 I I V I I I V^7 I I V I I I V^7 I
Poor 2-State:	A A A B B A A B B A B B A A B A A
Lexical Probs:	.8 .2 .8 .8 .2 .8 .8 .2 .8 .8 .2 .8 .8 .8 .2 .2 .8 = .0000055
Transition Probs:	.5 .5 .5 .5 .5 .5 .5 .5 .5 .5 .5 .5 .5 .5 .5 .5 = .0000153

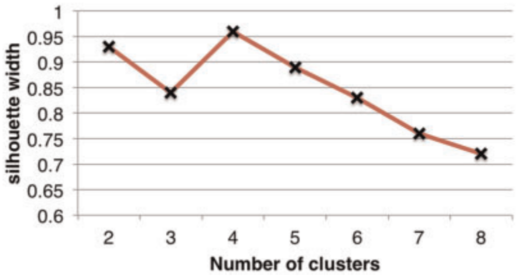
EXAMPLE A5. *Poor two-state HMM analyzing Toy Corpus no. 1*

Observations:	I V^7 I I V I I I V^7 I I V I I I V^7 I
Good 4-State #1:	A C A D B A A D C A D B A A D C A
Good 4-State #2:	A C A A B D A A C D A B D A A C D

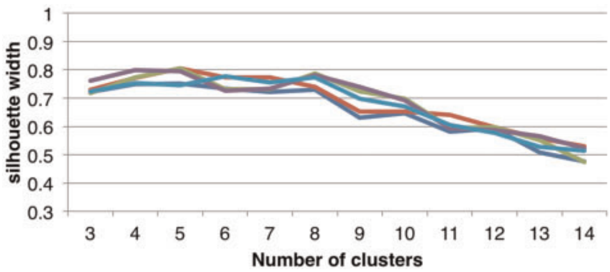
EXAMPLE A6. *Toy Corpus no. 1 analyzed by two Good four-State HMMs*

that sequence of assessments is used to produce a dissimilarity matrix, which is then in turn used for a *k*-means cluster analyses where *k* equals the number of hidden states used in the HMM. This process allows us to measure consistency in a situation where "the first hidden state" might mean something different for each training/test session. For instance, if we were modeling five runs of the HMM process using four hidden states, a possible vector might be [ADBCA], suggesting that the first HMM assigned the slice to state A, the second to state D, and so on. If this were a I chord, one might imagine that most other I chords would receive the same or similar analysis vector. Again, note that the state assignments are arbitrary: "A" might be "tonic" in one HMM, while "D" could be tonic in another. A "dominant" slice would then have a completely different vector, say, [BACBB]. A particular iii chord in the test corpus might have a vector more similar to tonic, while another might be more similar to the dominant. Each observation's vector—each chord's string of hidden states—within the test corpus is placed in a matrix, where the dissimilarity between each observation is calculated. The dissimilarity between [ADBCA] and [BACBB] would be five, since there is no overlap. The dissimilarity matrix now captures units of difference between how the models analyze every observation in the test corpus.[54] The cluster analysis is then performed using this matrix, dividing the observations into the number of hidden states within the constituent HMMs. The result is an

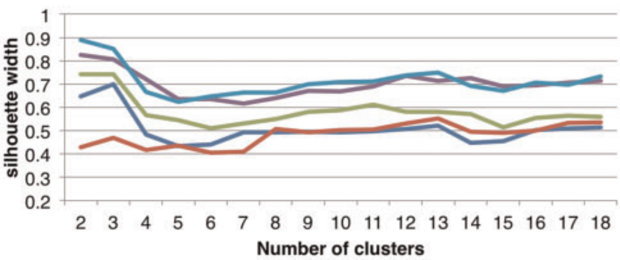
54 The current work squares the dissimilarities, simply to make the differences more pronounced, which in turn makes the data easier to work with.



EXAMPLE A7. *Silhouette widths of each cluster in the Kostka-Payne analysis*



EXAMPLE A8. *Silhouette widths of each cluster in the McGill-Billboard analysis*



EXAMPLE A9. *Silhouette widths of each cluster in the Bach-chorale analysis*

aggregated picture of which chords tend to be analyzed by which hidden states.

Example A7 shows the silhouette widths that result from the article's Kostka-Payne HMMs. The widths capture the amount of agreement—the tightness of the clustering—of the 300 models produced for each number of hidden states. Generally, the widths need to be read as relative to those around them since fewer clusters will tend to produce higher overall values. Therefore, we look for peaks within the contour: a peak will mean that the models improved their consistency compared to the predicted decline when adding clusters. (NB: there is no standard way of quantifying these peaks.) In Example A7, the widths peak at four clusters/states, indicating this provides the most consistent group of models, and therefore the preferred number of states for this corpus is four. Throughout the article, the models we diagram and report are summations of the agreement between the 300 individual models, and—since the clusters capture the emergent properties of the constituent HMMs—we create our final models by treating the clusters like composite functions. Since our clustering involves actual chords within our test set, we treat each cluster of chords as exemplifying its own function. The transitions between the clusters' constituent chords constitute the lexical probabilities, and the chords associated with each cluster become the emission probabilities.

Examples A8 and A9 show the silhouette widths of the five different test/training pairs for the popular music and Bach chorale studies. If peaks are reproduced over several trials, we consider the clustering robust. Note that Example A8 has consistent peaks at $k = 8$, while A9 has a consistent peak at $k = 3$, and a more subtle peak around $k = 13$ or 14.

APPENDIX B

PROBABILITY TABLES

	D/T$^-$	T	P-	P	End
D/T$^-$	0.4%	14.2%	0.4%	0.6%	0.1%
T	15.0%	0.0%	4.3%	5.4%	4.6%
P$^-$	0.0%	0.4%	2.5%	15.5%	0.4%
P	26.7%	0.3%	8.4%	0.3%	0.3%

EXAMPLE B1. *Transitions between hidden states of Kostka-Payne four-function model. Table sums to 100%, background colors are graded to reflect increasing probabilities.*

	Function			
chord root	D	T	P$^-$	P
1	0.00%	91.84%	0.00%	0.00%
#$\hat{1}$	0.48%	0.00%	10.71%	0.00%
$\hat{2}$	1.90%	0.00%	4.76%	65.00%
b$\hat{3}$ #2	0.00%	0.00%	0.00%	1.67%
$\hat{3}$	0.00%	2.04%	16.67%	0.00%
$\hat{4}$	9.52%	0.00%	9.52%	17.50%
#$\hat{4}$	1.43%	0.00%	1.19%	6.67%
$\hat{5}$	70.00%	0.00%	0.00%	0.00%
b$\hat{6}$ #5	1.43%	1.53%	3.57%	3.33%
$\hat{6}$	1.43%	1.53%	51.19%	0.00%
$\hat{7}$	12.38%	0.51%	1.19%	0.00%
Other	1.43%	2.55%	1.19%	5.83%

EXAMPLE B2. *Roots in Kostka-Payne associated with each hidden state (lexical probabilities). Columns sum to 100%.*

	P	Q	R	T	S	U	X	W	End
P	0.5%	2.9%	0.1%	0.0%	0.0%	0.1%	0.0%	0.0%	0.3%
Q	2.7%	0.9%	0.1%	0.1%	0.2%	0.0%	0.1%	0.0%	0.5%
R	0.0%	0.0%	0.0%	10.3%	2.8%	0.1%	0.0%	0.1%	0.6%
T	0.1%	0.1%	3.5%	0.3%	10.3%	10.0%	0.0%	0.0%	1.6%
S	0.1%	0.1%	9.6%	12.0%	0.1%	0.5%	0.0%	0.0%	1.1%
U	0.1%	0.1%	0.0%	1.8%	9.1%	0.4%	0.0%	0.0%	0.7%
X	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.6%	5.9%	0.6%
W	0.0%	0.1%	0.0%	0.1%	0.1%	0.0%	6.0%	1.6%	0.5%

EXAMPLE B3. *Transitions between hidden states of McGill-Billboard eight-function model. Table sums to 100%, background colors are graded to reflect increasing probabilities.*

	Functions							
chords	P	Q	R	T	S	U	X	W
I	26.8%	0.0%	0.0%	83.6%	0.0%	0.0%	8.2%	2.0%
IV	0.0%	2.3%	0.0%	0.0%	85.0%	0.0%	20.2%	0.0%
V	0.0%	10.5%	80.9%	0.0%	0.0%	30.2%	0.3%	0.2%
bVII	2.1%	0.9%	0.7%	3.2%	0.0%	22.0%	0.0%	23.3%
vi	2.1%	0.9%	0.0%	4.1%	0.4%	16.5%	0.0%	0.0%
$V^{sus\ 4}$	0.0%	0.0%	9.1%	0.6%	0.4%	6.8%	0.0%	0.0%
i	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	27.3%	0.0%
bVI	0.0%	0.0%	0.0%	0.6%	0.5%	0.0%	8.0%	10.6%
II	0.5%	0.0%	0.0%	0.3%	4.8%	2.2%	0.0%	0.0%
ii	1.1%	0.0%	0.0%	1.1%	2.8%	2.4%	1.4%	0.0%
iv	0.0%	1.4%	1.9%	0.2%	0.9%	0.0%	0.0%	8.9%
V^7	0.0%	11.4%	3.7%	0.0%	0.7%	0.3%	0.3%	0.0%
iii	0.0%	0.5%	0.1%	2.0%	0.0%	4.4%	0.0%	0.0%
i^7	0.0%	2.7%	0.0%	0.0%	0.0%	1.2%	0.0%	9.9%

EXAMPLE B4. *Chords in Kostka-Payne associated with each hidden state (lexical probabilities). Columns sum to 100%. Only chords that occur thirty times or more within the corpus are shown (>.05% of the corpus's chords).*

I no 3rd	0.0%	0.0%	0.0%	0.4%	0.0%	0.0%	11.1%	0.0%
iii^7	1.1%	1.4%	0.0%	0.8%	0.0%	4.6%	0.0%	0.0%
vi^7	3.2%	10.0%	0.0%	0.7%	0.3%	0.2%	0.3%	0.0%
v	0.0%	2.7%	0.1%	0.0%	0.0%	0.7%	3.4%	4.5%
ii^7	15.8%	0.5%	0.1%	0.0%	0.6%	0.0%	0.0%	0.2%
III	0.0%	0.0%	0.0%	0.5%	0.0%	1.0%	0.9%	5.7%
IV^9	0.0%	0.9%	0.0%	0.0%	3.2%	0.0%	0.0%	0.0%
v^7	3.2%	3.2%	0.4%	0.0%	0.0%	1.7%	0.3%	2.2%
I^9	0.0%	15.1%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%
$\{\hat{1}, \hat{2}, b\hat{7}\}$	0.0%	0.0%	0.0%	1.2%	0.0%	0.2%	0.0%	4.0%

EXAMPLE B4. *[Continued]*

	D	P	T	End
D	5.61%	0.91%	25.10%	0.49%
P	17.07%	22.61%	0.49%	1.27%
T	8.25%	16.73%	0.17%	1.30%

EXAMPLE B5. *Transitions between hidden states of Bach Chorale three-function model. Table sums to 100%, background colors are graded to reflect increasing probabilities.*

	Functions		
chords	D	P	T
I	0.0%	0.0%	73.1%
V	30.9%	5.7%	0.0%
IV	3.0%	15.8%	0.0%
V^7	21.1%	0.0%	0.0%
vi	2.0%	8.3%	5.4%
ii	0.5%	9.3%	0.8%
ii^7	1.6%	6.1%	0.0%
iii	1.9%	3.0%	4.0%
vii	7.6%	0.4%	0.0%
IV^7	0.0%	4.8%	0.0%
$\{\hat{1}, \hat{2}, \hat{5}\}$	1.2%	3.5%	0.0%
vi^7	0.0%	4.1%	0.2%
vii^7	5.0%	0.2%	0.0%
I^7	0.2%	3.3%	0.1%
$\{\hat{1}, \hat{4}, \hat{5}\}$	1.2%	1.7%	0.0%
II^7	0.0%	2.2%	0.0%
IV^9	0.0%	2.2%	0.0%
V/vi	0.5%	1.3%	0.6%
$\{\hat{1}, \hat{2}, \hat{4}\}$	1.1%	0.9%	0.0%
vii/vi	0.0%	1.7%	0.0%
$V^{7\text{ no 5th}}$	2.1%	0.0%	0.0%
V^7/IV	0.0%	1.5%	0.0%
V^7/vi	0.0%	1.2%	0.4%
$\{\hat{1}, \hat{5}, \hat{6}\}$	0.0%	0.5%	1.5%
$V^{\text{add 4}}$	0.7%	0.8%	0.0%
$\{\hat{1}, \hat{2}, \hat{6}\}$	0.0%	1.3%	0.0%
$\{\hat{2}, \hat{4}, \hat{5}\}$	1.6%	0.0%	0.0%
$\{\hat{1}, \hat{4}, \hat{6}, \hat{7}\}$	1.4%	0.2%	0.0%
V/iii	0.5%	0.8%	0.0%

EXAMPLE B6. *Chords in Bach Chorales associated with the three hidden states (lexical probabilities). Columns sum to 100%. Only chords that occur nineteen times or more within the corpus are shown (>.05% of the corpus's chords).*

	T+	R	Px	p	d	T	Tx	P	D	sx	Rx	D+	Dx	End
T+	0.2%	0.5%	0.3%	3.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.0%	0.0%	0.3%	0.0%
R	0.1%	0.0%	0.9%	0.4%	0.1%	0.6%	0.0%	0.2%	0.5%	0.2%	1.1%	0.0%	0.2%	0.1%
Px	0.1%	0.1%	0.2%	1.3%	0.6%	0.0%	0.0%	0.6%	0.6%	0.5%	0.0%	0.3%	0.1%	0.1%
p	0.0%	0.0%	2.6%	0.0%	3.6%	1.2%	0.0%	0.5%	1.0%	0.1%	0.0%	1.0%	0.1%	0.4%
d	0.0%	0.2%	0.1%	0.1%	0.0%	3.7%	0.1%	0.2%	0.0%	0.2%	0.1%	0.1%	0.0%	0.0%
T	3.4%	1.0%	0.1%	3.9%	0.1%	0.0%	3.2%	4.6%	3.2%	0.6%	0.3%	0.5%	0.1%	1.1%
Tx	0.1%	0.0%	0.0%	0.0%	0.0%	4.0%	0.2%	0.2%	0.1%	0.0%	0.0%	0.1%	0.0%	0.0%
P	0.0%	0.0%	0.0%	0.0%	0.1%	0.2%	0.0%	1.3%	5.5%	0.0%	0.1%	0.4%	0.1%	0.1%
D	0.3%	0.4%	0.0%	0.5%	0.0%	5.3%	0.2%	0.0%	0.1%	0.6%	0.2%	4.5%	1.8%	0.4%
sx	0.0%	0.3%	0.2%	0.4%	0.2%	0.2%	0.0%	0.2%	0.4%	2.9%	0.3%	0.3%	0.1%	0.4%
Rx	0.0%	1.6%	0.0%	0.5%	0.0%	0.1%	0.0%	0.0%	0.1%	0.1%	0.9%	0.0%	0.1%	0.3%
D+	0.0%	0.1%	0.0%	0.0%	0.0%	6.3%	0.9%	0.0%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%
Dx	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.5%	0.0%	0.2%	0.3%	1.1%	0.1%

EXAMPLE B7. *Transitions between hidden states of Bach Chorale thirteen-function model. Table sums to 100%, background colors are graded to reflect increasing probabilities.*

	functions												
chords	T+	R	Px	p	d	T	Tx	P	D	sx	Rx	D+	Dx
I	0.0%	0.0%	0.0%	0.0%	0.0%	87.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.6%
V	1.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.5%	0.0%	84.8%	0.0%	0.0%	0.0%	0.0%
IV	0.0%	0.0%	0.0%	72.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
V^7	0.0%	0.0%	0.0%	0.0%	1.0%	0.1%	0.0%	0.6%	0.2%	0.0%	0.0%	86.0%	0.0%
vi	0.5%	97.3%	4.3%	1.2%	0.0%	2.8%	2.5%	0.0%	0.0%	0.0%	0.0%	0.0%	1.2%
ii	0.0%	0.0%	5.9%	12.7%	1.0%	0.0%	0.0%	0.3%	0.0%	44.0%	0.0%	0.0%	0.0%
ii^7	0.0%	0.0%	0.0%	1.6%	7.9%	0.0%	0.0%	29.3%	0.0%	3.3%	0.0%	0.0%	0.0%
iii	7.2%	2.2%	4.8%	0.0%	0.5%	4.9%	3.0%	0.6%	5.0%	0.4%	4.5%	0.0%	1.2%
vii	3.6%	0.0%	0.0%	0.0%	39.9%	0.0%	6.0%	0.0%	0.2%	0.0%	0.0%	1.9%	0.0%
IV^7	0.0%	0.0%	23.7%	2.6%	0.0%	0.0%	0.0%	8.1%	0.0%	0.0%	0.0%	0.0%	0.0%
$\{\hat{1}, \hat{2}, \hat{5}\}$	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	4.0%	17.1%	1.5%	0.0%	0.0%	0.0%	2.3%
vi^7	0.0%	0.0%	27.4%	3.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	4.1%
vii^7	0.0%	0.0%	0.0%	0.0%	25.6%	0.0%	0.0%	0.0%	0.7%	0.0%	7.1%	0.0%	0.6%
I^7	28.4%	0.0%	0.0%	0.0%	0.0%	0.0%	1.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.6%
$\{\hat{1}, \hat{2}, \hat{4}, \hat{5}\}$	0.0%	0.0%	0.5%	0.0%	0.0%	0.0%	3.0%	15.0%	0.0%	0.0%	0.0%	0.3%	0.0%

EXAMPLE B8. *Chords in Bach Chorales associated with the thirteen hidden states (lexical probabilities). Columns sum to 100%. Only chords that occur nineteen times or more within the corpus are shown (>.05% of the corpus's chords).*

I^9	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	23.5%	0.0%	1.5%	0.0%	0.0%	0.0%	0.0%
$\mathrm{I}^{\text{add } 4}$	2.6%	0.0%	0.0%	0.0%	0.0%	0.0%	20.0%	0.0%	1.9%	0.0%	0.0%	0.0%	0.0%
iii^7	19.1%	0.0%	0.5%	0.0%	0.0%	1.3%	1.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.6%
$\{\hat{1}, \hat{4}, \hat{5}\}$	0.0%	0.0%	0.0%	0.0%	2.5%	0.0%	5.5%	7.8%	0.0%	0.0%	0.0%	1.3%	0.0%
II^7	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	3.7%	0.0%	0.0%	0.0%	0.0%	15.2%
IV^9	6.2%	0.0%	9.1%	0.9%	0.5%	0.0%	0.0%	0.9%	0.0%	0.0%	0.0%	0.0%	0.0%
V/vi	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	22.7%	0.0%	0.0%
$\{\hat{1}, \hat{2}, \hat{4}\}$	0.0%	0.0%	0.0%	2.8%	0.0%	0.0%	6.0%	2.2%	0.0%	0.0%	0.0%	0.0%	0.0%
vii/vi	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.6%	0.0%	0.0%	0.0%	0.0%	14.0%
$\mathrm{V}^{7 \text{ no 5th}}$	0.0%	0.0%	0.0%	0.0%	2.5%	0.0%	3.0%	0.0%	0.2%	0.0%	0.0%	4.8%	0.0%
V^7/IV	13.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
V^7/vi	2.1%	0.0%	3.8%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	7.0%
$\{\hat{1}, \hat{5}, \hat{6}\}$	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	15.6%	0.0%	0.0%
$\mathrm{V}^{\text{add } 4}$	5.7%	0.0%	0.0%	0.0%	0.0%	0.0%	2.5%	0.0%	0.0%	0.0%	0.0%	0.0%	3.5%
$\{\hat{1}, \hat{2}, \hat{6}\}$	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6.2%	0.0%	0.0%	0.6%	0.0%	0.6%
$\{\hat{2}, \hat{4}, \hat{5}\}$	0.0%	0.0%	0.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	11.7%
$\{\hat{1}, \hat{4}, \hat{6}, \hat{7}\}$	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	4.5%	0.3%	0.0%	0.8%	0.0%	2.9%	0.0%
V/iii	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	8.7%	0.0%	0.0%	0.0%
$\{\hat{2}, \hat{5}, \hat{6}\}$	1.0%	0.0%	0.0%	0.0%	9.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
v	1.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.5%	3.7%	0.0%	0.0%	0.0%

EXAMPLE B8. *[Continued]*

V^7/ii	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%	0.0%	0.0%	10.5%
II	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	7.9%	0.0%	0.0%	0.0%
$\{\hat{2}, \hat{3}, \hat{4}, \hat{6}\}$	0.0%	0.0%	2.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	5.4%	0.0%	0.0%	0.0%
vii/ii	0.0%	0.0%	0.0%	0.0%	0.0%	0.6%	0.0%	0.0%	0.0%	4.1%	0.0%	0.0%	0.0%

EXAMPLE B8. *[Continued]*

WORKS CITED

Aarden, Bret J. 2003. "Dynamic Melodic Expectancy." Ph.D. diss., Ohio State University.

Agmon, Eytan. 1995. "Functional Harmony Revisited: A Prototype-Theoretic Approach." *Music Theory Spectrum* 17 (2): 196–214.

———. 2013. *The Languages of Western Tonality*. Heidelberg: Springer Verlag.

Bello, Juan P., and Jeremy Pickens. 2005. "A Robust Mid-level Representation for Harmonic Content in Music Signals." In *Proceedings of the International Society for Music Information Retrieval.* 304–11. London.

Biamonte, Nicole. 2010. "Triadic Modal and Pentatonic Patterns in Rock Music." *Music Theory Spectrum* 32 (2): 95–110.

Burgoyne, J. Ashley. 2012. "Stochastic Processes and Database-Driven Musicology." Ph.D. diss., McGill University.

Burgoyne, J. Ashley, Jonathan, Wild, and Ichiro, Fujinaga. 2013. "Compositional Data Analysis of Harmonic Structures in Popular Music." In *Proceedings of the 4th International Conference on Mathematics and Computation in Music.* 52–63. Heidelberg: Springer.

Creel, Sarah C., Elissa L. Newport, and Richard N. Aslin. 2004. "Distant Melodies: Statistical Learning of Nonadjacent Dependencies in Tone Sequences." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30: 1119–30.

Cuthbert, Michael, and Christopher Ariza. 2011. "Music21: A Toolkit for Computer–Aided Musicology and Symbolic Music Data." In *Proceedings of the International Symposium on Music Information Retrieval.* 637–42.

Dahlhaus, Carl. 1968/1990. *Studies on the Origin of Harmonic Tonality.* Trans. Robert O. Gjerdingen. Princeton: Princeton University Press.

deClercq, Trevor, and David Temperley. 2011. "A Corpus Analysis of Rock Harmony." *Popular Music* 30 (1): 47–70.

Doll, Christopher. 2007. "Listening to Rock Harmony." Ph.D. diss., Columbia University.

Everett, Walter. 2004. "Making Sense of Rock's Tonal Systems." *Music Theory Online* 10.4.

Fétis, François-Joseph. 1844. *Traité complet de la théorie et de la pratique de l'harmonie*. Paris: Schlesinger.

Forte, Allen. 1962. *Tonal Harmony in Concept and Practice*. New Haven: Yale University Press.

Harrison, Daniel. 1994. *Harmonic Function in Chromatic Music: A Renewed Dualist Theory and an Account of Its Precedents*. Chicago: University of Chicago Press.

Hauptmann, Moritz. 1853 [1888]. *Die Natur der Harmonik und der Metrik*. Leipzig, Breitkopf und Härtel. Trans. W. Heathcote as *The Nature of Harmony and Metre*. London: S. Sonnenschein.

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

Jacoby, Nori, Naftali Tishby, and Dmitri Tymoczko. 2015. "An Information Theoretic Approach to Chord Categorization and Functional Harmony." *Journal of New Music Research* 44 (3): 219–44.

Jurafsky, Daniel, and James H. Martin. 2008. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 2nd edition. Upper Saddle River, NJ: Prentice Hall.

Klumpenhouwer, Henry. 2007. "Dualist Tonal Space and Transformation in Nineteenth-Century Musical Thought." In *The Cambridge History of Western Music Theory*, 5th edition. Ed. Thomas Christensen. 456–76. Cambridge: Cambridge University Press.

Kopp, David. 2002. *Chromatic Transformations in Nineteenth-Century Music*. Cambridge: Cambridge University Press.

Kostka, Stefan, and Dorothy Payne. 2012. *Tonal Harmony with an Introduction to Twentieth-Century Music*, 4th edition. New York: McGraw-Hill.

Krumhansl, Carol L. 1988. "Tonal and Harmonic Hierarchies." In *Harmony and Tonality.* Ed. J. Sundberg. Stockholm: Royal Swedish Academy of Music.

———. 1990. *The Cognitive Foundations of Musical Pitch*. Oxford: Oxford University Press.

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

Lester, Joel. 1992. *Compositional Theory in the Eighteenth Century*. Cambridge, MA: Harvard University Press.

Loui, Psyche, David Wessel, and Carla L. Hudson Kam. 2010. "Humans Rapidly Learn Grammatical Structure in a New Musical Scale." *Music Perception* 27: 377–88.

Loui, Psyche. 2012. "Learning and Liking of Melody and Harmony: Further Studies in Artificial Grammar Learning." *Topics in Cognitive Science* 4: 1–14.

Mavromatis, Panayotis. 2005. "A Hidden Markov Model of Melody Production in Greek Church Chant." *Computing in Musicology* 14: 93–112.

———. 2009. "HMM Analysis of Musical Structure: Identification of Hidden Variables through Topology-Sensitive Model Selection." In *Communications in Computer and Information Science: Mathematics and Computation in Music*. Eds. Ellen Chew, Adrian Childs, and Ching-Hua Chuan. 205–217. Heidelberg: Springer Verlag.

———. 2012. "Exploring the Rhythm of the Palestrina Style: A Case Study in Probabilistic Grammar Induction." *Journal of Music Theory* 56: 169–223.

McHose, Allen Irvine. 1947. *The Contrapuntal Harmonic Technique of the 18th Century*. New York: F. S. Crofts and Co.

Momigny, Jérôme-Joseph. 1806. *Cours complet d'harmonie et de composition d'après une théorie neuve et générale de la musique*. Paris: Bailleul.

Moore, Allan.1992. "Patterns of Harmony." *Popular Music* 11 (1): 73–106.

Nobile, Drew. 2014. "A Structural Approach to the Analysis of Rock Music." Ph.D. diss., CUNY Graduate Center.

Quinn, Ian. 2010. "What's 'Key for Key': A Theoretically Naive Key–Finding Model for Bach Chorales." *Zeitschrift der Gesellschaft für Musiktheorie*7(ii): 151–63.

———. 2017. "Tonal Harmony." In *The Oxford Handbook of Concepts in Music Theory*. Ed. Alexander Rehding and Steven Rings. New York: Oxford University Press.

Quinn, Ian, and Panayotis Mavromatis. 2011. "Voice Leading and Harmonic Function in Two Chorale Corpora." In *Mathematics and Computation in Music*. Ed. Carlos Agon. 230–40. Heidelberg: Springer.

Rameau, Jean-Philippe. 1737. *Génération harmonique, ou Traité de musique théorique et pratique*. Paris: Jean-Baptiste-Christophe Ballard.

Raphael, Christopher, and Joshua Stoddard. 2004. "Functional Analysis Using Probabilistic Models." *Computer Music Journal* 28 (3): 45–52.

Rehding, Alexander. 2011. "Tonality between Rule and Repertory; Or, Riemann's Functions—Beethoven's Function." *Music Theory Spectrum* 33 (2): 109–23.

Riemann, Hugo. 1893. *Vereinfachte Harmonielehre, oder die Lehre von den tonalen Funktionen der Akkorde*. London: Augener.

———. 1900–1901. *Präludien und Studien III.* Leipzig: Hermann Seemann.

Riepel, Joseph. 1755. *Grundregeln zur Tonordnung insgemein: Abermal Durchgehends mit musicalischen Exempeln abgefaßt und Gespräch-weise vorgetragen*. Frankfurt: Krippner.

Saffran, Jenny R., Elizabeth K. Johnson, Richard N. Aslin, and Elissa L. Newport. 1999. "Statistical Learning of Tone Sequences by Human Infants and Adults." *Cognition* 70: 27–52.

Saffran, Jenny R., Karelyn Reeck, Aimee Niebuhr, and Diana Wilson. 2005. "Changing the Tune: The Structure of the Input Affects Infants' Use of Absolute and Relative Pitch." *Developmental Science* 8 (1): 1–7.

Smith, Charles. 1981. "Prolongations and Progressions as Musical Syntax." In *Music Theory: Special Topics.* Ed. Richmond Browne. 139–74. New York: Academic Press.

Stein, Deborah. 1983. "The Expansion of the Subdominant in the Late Nineteenth Century." *Journal of Music Theory* 27 (2): 153–80.

Stephenson, Kenneth. 2002. *What to Listen for in Rock: A Stylistic Analysis.* New Haven, CT: Yale University Press.

Swinden, Kevin J. 2005. "When Functions Collide: Aspects of Plural Function in Chromatic Music." *Music Theory Spectrum* 27 (2): 249–82.

Temperley, David. 2009. "A Statistical Analysis of Tonal Harmony." Accessed September 2, 2018. http://davidtemperley.com/kp-stats/.

Tymoczko, Dmitri. 2003. "Function Theories: A Statistical Approach." *Musurgia* 10 (3–4): 35–64.

White, Christopher W. 2013. "An Alphabet-Reduction Algorithm for Chordal *n*-grams." In *Proceedings of the 4th International Conference on Mathematics and Computation in Music*. Eds. Jason Yust, Jonathan Wild, and John Ashley Burgoyne. 201–12. Heidelberg: Springer.

White, Christopher W., and Ian Quinn. 2016. "The Yale-Classical Archives Corpus." *Empirical Musicology Review* 11 (1): 50–58.

Wittig, Susan. 1977. "The Computer and the Concept of Text." *Computers and the Humanities.* 11 (4): 211–15.

Music Theory Spectrum, Vol. 40, Issue 2, pp. 314–37, ISSN 0195-6167, electronic ISSN 1533-8339.
DOI: 10.1093/mts/mty021