6   Computational Chord-Root Identification in Symbolic Musical Data: Rationale, Methods, and Applications

Craig Stuart Sapp
Center for the History and Analysis of Recorded Music
Royal Holloway College
University of London
Egham, Surrey TW20 0EX, UK
craig.sapp@rhul.ac.uk

Abstract

Our approach to tonal harmonic analysis, developed after evaluating a number of other approaches, involves the analysis of pitch-class, rhythmic and metric context, melodic content, tonality, and form. Datasets examined are encoded in **kern. Particular emphasis is given to assessing problems of variable texture, in which chordal information is ambiguous or incomplete. Evaluation of passing tones, inner voices, and other minutiae of musical scores which interfere with simpler chordal assessments are among the topics discussed. Recent applications of the evaluation procedures are also described.

6.1   Present Goals

There are numerous motivations for developing automatically generated harmonic analyses of music. First, automatic analysis can relieve the tediousness of manual analysis. More important, it is a useful step towards providing computers with a deeper understanding of music necessary for applications such as automated performance and generative composition. Musical scores are intended for human interpretation, and computers' lack of musical insight can make their literal interpretations difficult to listen to. For example, a computer "performance" of a musical score from a quantized MIDI file can be quite mechanical: while a performer may slow down at the ends of phrases almost instinctively, a computer has no intrinsic concept of phrasing or cadences and will happily barrel through the score at a constant pace.

The underlying harmonic structure of tonal music contributes significantly to an interpretation. Providing this information to a computer is a large step towards generating engaging musical performances or compositions. In music query, harmonic analysis can be used to identify related segments of music even when the surface texture is different from one instance to the next. Harmonic analysis also reveals the templates of many forms in Western music, and in this sense is essential to structural analysis of complete works such as sonatas.

At the core of harmonic analysis lies the correct identification of chord roots. Once the root is determined, it is fairly easy to determine the quality of a chord (i.e., is it major, minor, augmented, diminished, etc.). Once the chord root and quality have been determined, it is not much more difficult to assign a functional label (e.g., tonic, dominant, subdominant, etc.) if the music is tonal. The identification of chords in a continuous stream of symbolic data representing a musical score is, however, complicated by a conundrum. In order to separate chords and identify chord roots, it is necessary to identify non-harmonic tones. In order to determine whether tones are non-harmonic, one needs to identify a chord root against which to compare them. In real musical textures, as opposed to textbook extrapolations of them, chords have the further properties that they can occur in a vast number of configurations and can contain an unspecified number of members lying beyond the normal root-third-fifth span. These factors make the analysis of tonal harmony in symbolic data of real music far from trivial.

My aim here is to present a basic methodology for identifying the roots of chords in symbolic score data encoded in the Humdrum kern format. (Almost 1,000 works encoded in this format can now be found at the websites *kern.ccarh.org* or *kern.humdrum.org*.) The primary approach tests four perceptual contexts of each candidate chord. These perceptual contexts, in order from local to global information about the notes in a chord, are:

PITCH CONTEXT: assessed by arranging chord tones into stacks of thirds and calculating the most compact arrangements.

RHYTHMIC AND METRIC CONTEXTS: weighting pitch information by duration and metric information to refine the calculation of likely chord roots.

MELODIC CONTEXT: incorporating information about melodic resolutions between chord and non-chord tones to further refine chord root identification.

KEY CONTEXT: ambiguous identifications of root tones in the previous contexts can often be resolved by considering the musical key in which the chord is situated. A simple example of this would be a fully diminished seventh chord which has the same quality in any inversion, but its pitch-classes are spelled according to the following chord to which it resolves.

The overall process starts with an initial estimation of the chord root based on limited information from pitch alone, then refines the measurements with additional information from a gradually wider connection to other notes and chords in the surrounding music. After these operations, either a cognitive decision on root identification occurs, or an ambiguity in the harmony can be identified.

6.2   Recent Approaches

Among the many recent writings on the analysis of tonal music, my approach has been stimulated to the greatest degree by the theses of John Maxwell (1984), who approached analysis from the perspective of artificial intelligence; David Temperley (1996), who approached it from the linked perspectives of perception and cognition; and Elaine Chew (2000), who presented a mathematical model of tonal organization.

Maxwell's analysis is based on the joining of individual sonorities into chords. He defines a sonority as a set of harmonic (and possibly non-harmonic) tones sounding at the same time. Dissonant sonorities are linked to adjacent consonant sonorities based on the resolution of non-harmonic tones. His was an expert system based on 55 rules.[1] As in the example cited in the note, the rules consist predominantly of Boolean expressions. Maxwell's approach works better for some musical genres than for others. It is particularly useful for music which is both contrapuntal and tonal. The system was subsequently expanded into a LISP-based program for harmonic analysis developed by Taube (1999). Some weakness that occur in this system concern (1) the use of the sonority as the basic unit (it does not work well for textures which are elaborate), and (2) the fact that the rules are applied serially rather than in parallel. Since rules must be applied in sequence, erroneous output from a low-level rule can be magnified as it is propagated onto later applications of higher-level rules.

Temperley's thesis (1996) and its extensions and enhancements in his first book (2001) represent an adaptation of the generative theory of tonal music (GTTM) of Lerdahl and Jackendoff (1983). Essentially, Temperley shapes the preference rules of GTTM to fit a digital environment. He uses this digital environment to produce

many of the findings reported in his second book (2007), which, however, offers an alternative (Bayesian) model for its analytical routines.

In Temperley (2001) harmonic analyses start from the lowest level of a symbolic score. The musical data in his *Melisma Music Analyzer*[2] is essentially MIDI data: it includes the start and stop times of a note in milliseconds and the MIDI-key number of the note (i.e., there are no pitch spellings).[3] The data is filtered through several programs before a final tonal harmonic analysis is assigned. There are three steps in the process: (1) identification of meter and beat positions; (2) identification of enharmonic spellings of notes and the generation of a root analysis; and (3) assignment of key and functional harmonic labels.[4] Temperley's system is more robust than that of Maxwell primarily because individual rules can be overturned by other rules with higher priority if circumstances warrant it. The system analyzes Classical-era music very well and does surprisingly well when dealing with Romantic-era music.

Chew's geometrical model of pitches (2000), used to identify the key of a musical excerpt, has certain conceptual similarities to my work since both utilize spatial relationships between notes in the calculation of harmonies. Her spiral-array model features the placement of the line of fifths (e.g., …, Bb, F, C, G, D, …) on a coil. The spatial arrangements on the spiral express harmonic interval strength, since closely related harmonic pitches are close to each other. Pitches are combined into chords by delineating triangular regions between notes. The centers of these triangles connect with other chord centers, which then give rise to key regions. Chew's system is optimized for identifying the key of a region of music. The system can recognize chords, but it is not always accurate in identifying chords other than triads, since chordal identification is geared towards relationships of three-note groups.

6.3   Pitch Assessment of Chords in Variable Textures

If musical texture consisted of nothing but chords, one of the most daunting problems of harmonic detection would not exist. Textbook examples of harmonic progressions depend on the ability of the human brain to extrapolate chordal identity from a changing tonal context. The tonality of piece of music, a fundamental principle of organization over the eighteenth and nineteenth centuries, is itself a cognitive extrapolation. At any given moment, a listener or a performer may not be conscious of the local key. Yet analyses invariably demonstrate the malleability of the key. In fact, much of the sport of composition in a fixed tonality involves the manipulation of key. In Figure 6.1, we see a very simple passage of four-voice harmony in which the alto voice moves independently of the other voices.

Figure 6.1. *Independent voice movement in four-part harmony.*

To describe it in chordal terminology requires five symbols, even though many people think of it as containing only three chords. Figure 6.2 shows the representation of the four voices of this passage in **kern, used with the Humdrum Toolkit. To these four voices are appended two further spines. One shows the presumed root of the chord at each event, while the other translates these into Roman numerals.

bass	tenor	alto	soprano	root	analysis
**kern	**kern	**kern	**kern	**root	**harm
=4	=4	=4	=4	=4	=4
2GG;	2B-;	2d;	2g;	2GG;	v;
4GG	4B-	8d	4g	4GG	v
.	.	4e-	.	.	.
4FF#	4c	.	4a	8FF#	viioD7/v
.	.	[8d	.	8DD	V7b/v
=5	=5	=5	=5	=5	=5
*-	*-	*-	*-	*-	*-

Figure 6.2. *Humdrum **kern representation of the four voices in Figure 6.1.*

6.3.1 Root Estimation by Interval Compactness

As a first and most important approximation in identifying the root of a chord, candidate chords are evaluated by a simple pitch-only harmonic model. The pitch-classes of a chord are arranged to form a sequence of thirds above potential root pitch-classes. Arrangements of pitches into these "stacks of thirds" which form more compact chords can then be used to identify the possible root of a chord. At this stage, no distinction is made between chordal and non-chordal notes, and all pitches are presumed to be chord tones.

In the following section, a basic root-finding algorithm is presented starting with diatonic intervals. Chords are built up from the intervals of the prime, third, fifth, seventh, ninth, eleventh, and thirteenth (the last three are not usually necessary for simple harmonic analysis, but they are valuable in modulatory contexts and essential for later tonal repertories such as late nineteenth-century classical music and jazz). These interval classes can be assigned numerical values used to calculate good candidates for the root of a chord. The pitch-only algorithm is then generalized to full chromatic intervals for practical use on real musical examples, and several spatial arrangements of pitches to generate interval-class weightings are examined.

ROOT IDENTIFICATION BY STACKS OF THIRDS

For any given collection of pitches, a series of possible chord roots can be evaluated by arranging the pitch-classes of the notes into stacks of thirds above each possible root. Each increment of a third above the root can be assigned a numeric value such as 1. For example, if the test root were "C", then the "C" pitch-classes in a musical extract would be assigned the value 0, and "E" would be assigned the value "1" since it is an interval of a third above "C"; "G" would then be assigned the value 2, "B" assigned "3", and so on up to the pitch-class "A" which would be assigned the value 6, since it takes 6 thirds to reach "A" from "C" in a stack of thirds rising from the root "C". Figure 6.3 shows the arrangement of intervals and their weights when the chord root starts on "C" and on "D". The stack of thirds can also be thought of as interval classes which have values associated with them. For example, a unison with a test-root has the value 0, a third is 1, fifth = 2, seventh = 3, ninth (second) = 4, eleventh (fourth) = 5, and thirteenth (sixth) = 6.
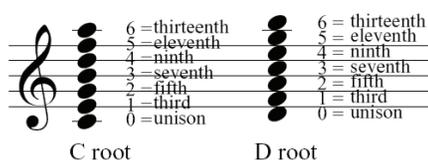
Figure 6.3. *Representative values of diatonic interval classes for roots on C and D.*

Once the value of each pitch in the musical excerpt has been assigned a particular test root, all of these interval-class values can be added to generate a score representing the number of thirds represented in the chord. This summation can then be carried out using other test roots, such as for "D" where the "E" pitch-class value changes from 1 to 4 since "E" is four thirds, or a ninth above "D" in a stack of thirds. After all summations have been calculated for each test root, the lowest summation gives the best fit for a root on the particular pitch-class representing one of the test roots.

test root:	C	D	E	F	G	A	B
score:	9	11	6	15	3	12	7
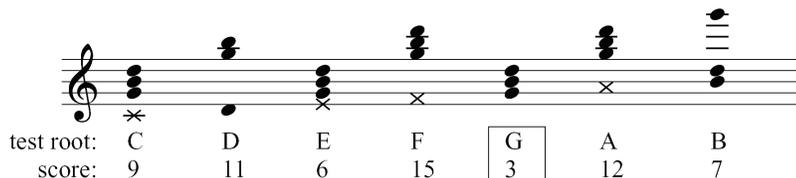
Figure 6.4. *Chordal compactness scores for the notes B, D ,G against seven possible roots. "X" represents the hypothetical root if no actual chord tone is present at the unison with the root pitch-class.*

Figure 6.4 illustrates the steps for finding the root of the notes "B", "D", and "G". First a summation of thirds is done when considering "C" as the root. In this case, "G" has a value of 2 since it is two thirds above "C". "B" and "D" respectively are assigned the values 3 and 4. Thus the summation of thirds when the root is on "C" in this example is 2+3+4 = 9. Considering "D" as the test root will yield a summation of

0+5+6 = 11 since "D" is located at the test-root position, and "G"/"B" are 5/6 units above a root on "D". Once all of the summations are made for each test root, the low score of 3 is noted for the test root "G", which is then selected as the best choice as a root for the pitches "B", "D", and "G".

Figure 6.5 illustrates a more complex situation where the notes in the chord are spread out over time, containing both chord and non-chord tones as well as different durations and metrical positions. In this case, the pitch-class content of the chord is "A", "B"x2, "C", "D" and "G". Most musicians would hear a root on "G" in this example, even though the "D" has the longest duration, and the "G" is delayed. Nonetheless, the lowest score when calculating the number of thirds found in the chord occurs on the test root "G" as demonstrated by the interval weight summations found on the right side of Figure 6.5.

test root:	C	D	E	F	G	A	B
score:	18	22	19	23	13	17	14
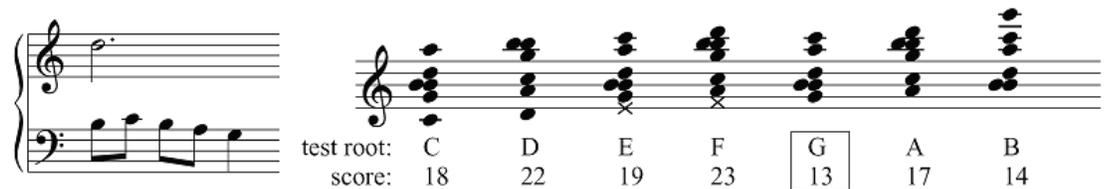
Figure 6.5. *Musical excerpt (left) containing changing and passing tones, and an assessment of diatonic test roots (right).*

Mathematically, this process of identifying the best fit for a root pitch-class is given in the following equation:

$$R = \min_p \sum_n I_{n,p}$$

(1)

where R is best measured root, p is the set of test roots (pitch-classes), n is an enumeration of all notes in the musical example, and $I_{n,p}$ is the numeric weight for the interval of note n given the test root's pitch-class p. Sigma represents a summation operation in mathematics, and "min" means do separate summations for all values of p (test root pitch-classes), and select the pitch-class which yields the lowest summation result. Notice that the lowest score occurs when the root is "G", however a root on "B" has a nearly identical score: If there were three "B" notes in the music, then the scores would be tied between the possible roots on "G" and "B".

EXTENDED MUSICAL EXAMPLE

Figure 6.6 shows an excerpt of actual music along with a harmonic analysis underneath as well as the pitch-class for the root of each chord. The pitch-classes of the notes in each chord region are given in the table shown in Figure 6.7. Note that rhythmic information is not being considered at this stage, and only the note pitches are being used to identify the root of the chords.

C:	V	V₇	vi	vii°	I	V₇/V	V₆	iii	vi₇	ii	ii	vii°₆	V₇	I	ii₇	vii⌀₇/IV

C: V V$_7$ vi vii° I $\underset{V}{V_7}$ V$_6$ iii vi$_7$ ii ii vii°$_6$ V$_7$ I ii$_7$ $\underset{IV}{vii^{\o}_7}$

root: G G A B C D G E A D D B G C D E

chord: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Figure 6.6. *J. S. Bach, Well-Tempered Clavier, Book I, Fugue 1, Bars 4 and 5.*

chord	pitch set	chord	pitch set	chord	pitch set	chord	pitch set
1	B G G	5	C C D E E	9	A C G	13	B F F G
2	A B F G	6	D F♯ C	10	A D E F	14	C C E G G
3	A B C E	7	B G	11	B C D E F	15	C D D E F
4	B C D D	8	B E G	12	A B D F	16	B♭ D E F G

Figure 6.7. *Pitch-class content for each chord region in Figure 6.6.*

For each of the chord pitch-class sets found in Figure 6.7, the root identification method of Equation 1 matches to the same root identified manually in Figure 6.6.

6.3.2 Chromatic Interval-Class Generalization

The diatonic interval space which is used to generate triadic distances from the root to other pitch-classes found in the chord can be generalized to a chromatic space which is more useful for practical application in music written in any musical key. In the construction of a triad in tonal music, there are two types of intervals which occur in the stacks of thirds that build a chord: the major third and the minor third. Figure 6.8 illustrates the options for placing a chordal note above "C". In diatonic space, the note above "C" is an "E" with any accidentals applied to the notes being ignored. In a chromatic context, there are exactly two options: either E natural which forms a major third above "C", or an E-flat which forms a minor third above "C".
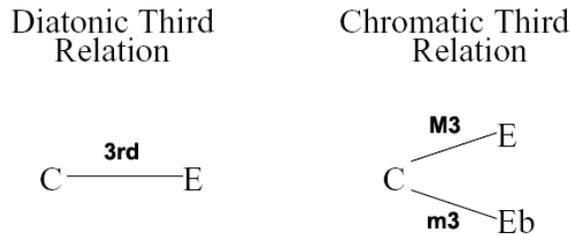
Figure 6.8. *Relationship between diatonic and chromatic thirds.*

To build a full triad, the process can also be extended starting on E and E-flat. Figure 6.9 displays how the second level of thirds in a chord with a root on C natural consists of three options: G-sharp, G-natural, and G-flat. This configuration can be used to describe the four possible theoretical triads in music theory: major triad, minor triad, diminished triad, and augmented triad.

Figure 6.9. *Major/minor third pathways visualized geometrically, and resulting triads.*

The process of stacking chromatic thirds above the root can continue further until all chromatic intervals are incorporated. Figure 6.10 shows this chromatic stack-of-thirds which forms the same pitch organization as Euler's *Tonnetz*.

Figure 6.10. *Diatonic and chromatic interval-class arrangements with an example root on "C".*

The chromatic and diatonic stacks of thirds can be related to each other by controlling the angle between the major and minor third in the second dimension. When the angle between the major and minor third is set to zero, as illustrated in Figure 6.11, the chromatic and diatonic spaces are equivalent (or nearly so when ignoring more exotic intervals such as the augmented unison). As the angle increases to 60 degrees, the ratio of the perfect fifth to the major/minor third changes from 2:1 to 1:1.

Figure 6.11. *Parametric angle for chromatic interval space (equivalent to diatonic space when the angle is set to 0).*

To evaluate the relative usefulness of diatonic and various forms of the chromatic arrangement of pitches, the root-finding algorithm of Equation 1 can be used to measure the root-identification error rate given interval weightings derived from various pitch configurations. Figure 6.12 shows the measured error rates using a set of 3,200 chords from 50 Bach chorales. Notice that the error rate remains steady between 11 and 12 percent in the range, but when the perfect fifth interval weight becomes less than that of the major/minor third, the error rate suddenly jumps. As the ratio of the fifth to the third further decreases as the parametric angle increases above 60 degrees, the error rate quickly increases.
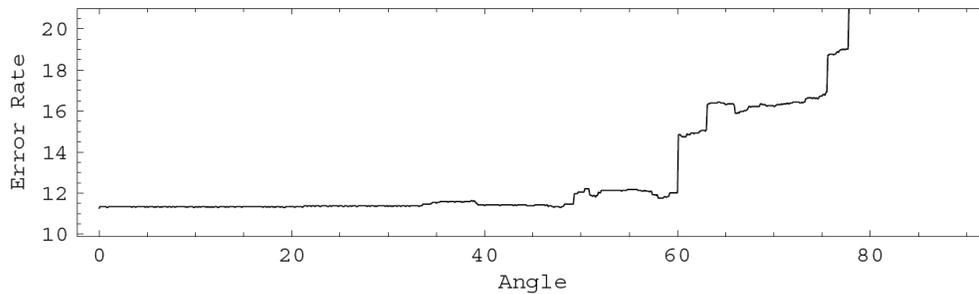
Figure 6.12. *Identification error rate in a test collection of chords based on the chromatic angle space.*

6.3.3   Other Pitch/Interval Spaces

The diatonic arrangements of thirds and the Euler *Tonnetz* configuration are low-dimensional pitch spaces which can be used to calculate interval-class weights. However, other arrangements can also be examined.

The *Tonnetz* arrangement of the pitch-classes does not necessarily represent the best arrangement of pitches to calculate interval weights. Also, not all interval classes have equal importance when finding the chord root by intervallic compactness.

Figure 6.13 displays a plot showing the sensitivity of various interval classes in the identification of the correct root. The horizontal axis represents the numeric value of each chromatic interval class. The horizontal scaling and offset is arbitrary and will not affect the root identification. Typically the unison (labeled C) is given a value of zero, and the perfect fifth (labeled G) is given a value of one when comparing various pitch-class configurations. The hollow and filled circles just distinguish the various pitch/interval classes. The interesting part of the plot is the horizontal lines attached to each circle. These lines indicate the amount variability that can be applied to a single interval weighting, while keeping all other weightings constant, yet still yielding the same or better identification error rate.

For example, the horizontal line that extents to the left of the circle labeled "F" represents the range of interval weights for a perfect fourth which will generate the same or better identification rate for the chord test set. If the interval weight for P4 (a perfect fourth) is reduced by up to half of its original value, the identification error rate will be as good or better than if the original weight were used. In other words, the presence of a perfect fourth is not as damaging for root identification as predicted by the basic *Tonnetz* pitch layout. Keep in mind, however, that this analysis applies to a specific test set of chords, and results may change if other test sets are used.

Nonetheless, notice that there are two basic categories of interval classes in Figure 6.11: interval weights which can be varied a lot without changing the results, and interval weights which can only be changed a little before worsening the results. The former are mostly interval classes which are very rare (above the root), and not usu-
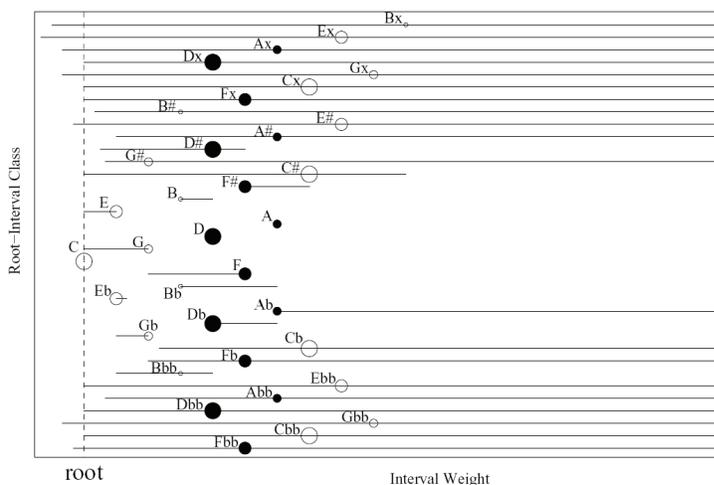
Figure 6.13. *Interval weight sensitivity for chromatic interval-class arrangements (angle = 60 degrees).*

ally found in chords. These interval weights can be set to just about any value, and their error rate will remain the same or better than the original weight: all of the (doubly) diminished intervals except the d5 and d7, as well as all of the augmented intervals except for the A4 and A2.

The fifteen other intervals have various tendencies; M2 and M6 weights cannot be changed without raising the error rate. The weights for M3, P5, d5 and P4 can all be lowered, yet keep the error rate the same or better. The weights for minor intervals (m2, m3, m6, m7), M7, and A4 can all be increased while keeping the error rate constant or reducing it.

Using the tendencies of the interval-class weights in the *Tonnetz* to either increase or decrease, other arrangements of pitches/intervals can be designed and evaluated. Figure 6.14 shows a set of interval-class weights which yields a 50% decrease in the error rate of the test set of chords. Most intervals in the diatonic area have become very sensitive and cannot be changed without increasing the error rate.
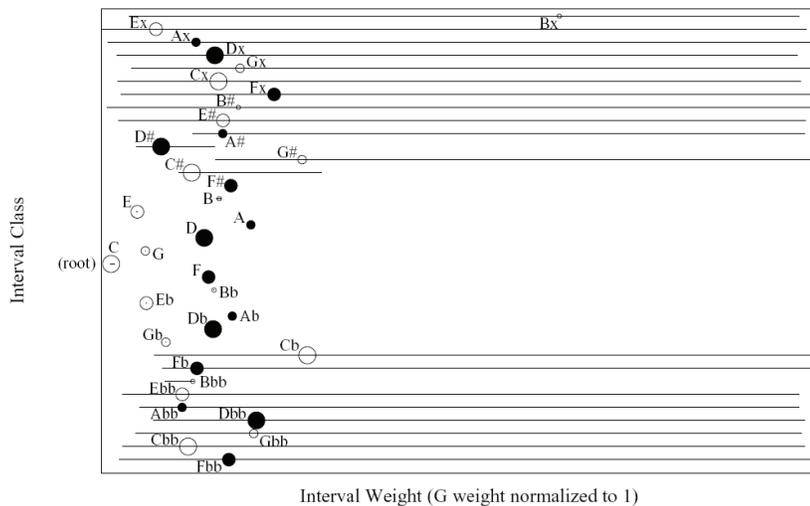
Figure 6.14. Monte Carlo and simplex optimization of the error rate.

The interval weights given in Figure 6.14 were generated by Monte Carlo methods similar to simulated annealing. First, all interval-class weights were set to zero. All chords were analyzed with this initial set of weights which will not find the correct roots very well. Then a random value is added to each interval weight (such as a number between 0 and 1), and the test set is then analyzed for errors using the new weights. The previous step is repeated a large number of times (such as a million times), adding different random values to the original zero-interval weights. Then the configuration which gives the best error rate is set as the new default position from which a new million random searches are done. Each time the default position changes to a new set of values, the random value range decreases (such as 5% after each move).

After the Monte Carlo method reaches the limit of its optimization, the downhill simplex method is used to search for the nearest local minimum in the error rate by slightly adjusting the ending interval weights derived from simulated annealing. Thus, the horizontal lines show the equivalent error-rate range, since the weights have been adjusted to the local minimum of the error function.

6.4 Chordal/Non-Chordal Differentiation

6.4.1 Rhythmic Contexts of Non-Harmonic Tones

Pitch measures alone may not be sufficient to accurately identify all types of chords. In works of the eighteenth and nineteenth centuries, non-harmonic tones almost always resolve by step to a chord tone, or depart by step from one. The evaluation of these tones cannot be made from pitch-class alone, because various ambiguities arise. It benefits from assessments of meter and duration, because accented non-harmonic tones contradict rhythmic scaling preferences. For example, it is possible for two chords to have the same pitch-classes, yet different roots (Example 6.15).

Figure 6.15. *Three chords with different roots but similar pitch-classes (each collection contains the pitches C, E, and G).*

The two primary rhythmic measurements available from a musical score are (1) the duration and (2) the metric position of a note. To create a measure of rhythmic scaling, we combined measures for scaling both duration (δ) and meter (λ). With regard to duration, the whole note, half note, quarter note, eighth note, and sixteenth note are scored 4, 2, 1, 0.5, 0.25 in binary meter contexts. With regard to metrical position, they are scored 2, 1, 0, -1, -2. The combination produces a binomial expression. Some examples are given in Figure 6.16.

Figure 6.16. *Binomial expressions for discrimination of a series of notes of (logically) identical duration (and pitch) with different metric placements in common time.*

These rhythmic and metrical qualities may be incorporated into the (pitch-based) chordal-compactness algorithm by scaling the root-interval weighting of the pitch by its duration and/or metric position. Equation 2 shows the duration-scaling formula

$$\text{root} \;=\; \min_{j} \sum D_n^{\delta} I_{jn}$$

(2)

where D_n is the duration of the nth note in a chord, and delta (δ) is a scaling factor which indicates the relative importance of duration to the identification of the chord root. (Note that if δ is set to 0, then the duration-scaling of pitch information is removed.) Similarly, Equation 3 shows the metrical scaling formula

$$\text{root} \;=\; \min_{j} \sum L_n^{\lambda} I_{jn}$$

(3)

where L_n is the metrical weight of the nth note and lambda (λ) is the scaling factor.

For passages with many passing tones, such as the one shown in Example 6.9, the effects of rhythmic and metrical scaling on root identification are dramatic. On the basis of pitch alone, the best root estimation for Example 6.17 indicates C as the root in all three cases.

test root:	C	E	G	measured root	actual root
chord 1	11	12	20	C	C
chord 2	13	14	15	C	G
chord 3	12	14	14	C	E

Figure 6.17. *Using only pitch as a basis for root assessment, the estimated and actual roots differ in two cases out of three.*

When rhythmic and metric evaluation are added, with $\delta = 0.25$ and $\lambda = 0.25$, the estimates change significantly (Figure 6.18).

test root:	C	E	G	measured root	actual root
chord 1	10.38	12.09	23.69	C	C
chord 2	16.10	14.43	12.76	G	G
chord 3	13.24	11.57	13.27	E	E

Figure 6.18. *Root assessment in which pitch information and rhythmic information are combined.*

6.4.2   Melodic Contexts of Non-Harmonic Tones

PASSING TONES

From these evaluations, we may proceed to differentiate between non-harmonic tones which conform to our rhythmic-scaling model and those which do not. In the first group we include *unaccented* passing tones, neighboring tones, melodic anticipations, and escape tones (Figure 6.19a).

Unaccented Passing Tone   Neighboring Tone   Anticipation   Escape Tone

S± S±   S± S∓   S± E   S± L∓

More problematical are *accented* passing tones, melodic suspensions (shown here without a tie from the preceding note), retardations, and appoggiaturas (Figure 6.19b).

Accented Passing Tone   Suspension   Retardation   Appoggiatura

S± S±   E S−   E S+   L± S∓

Figures 6.19a and 6.19b. *Selected categories of non-harmonic tones which do not conform to our rhythmic-scaling model. The nature of pitch motion is indicated by letter: E = equal (no pitch motion), S = step (+ = up, − = down), and L = leap.*

One additional category consists of what we call hybrid figures (Figure 6.20).

Changing Tones   Pedal Tone

S±   S±   E   E

E
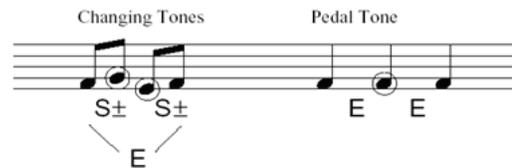
Figure 6.20. *Melodic figures containing passing tones with mixed accentuation features.*

Unaccented passing tones can be distinguished by rhythmic scaling, but accented ones and mixed figures, such as those shown in Figure 6.20, cannot. Metrical scaling is of particular value in evaluating them. Excluding only the pedal tone in Figure 6.20, all the passing tones shown in Figures 6.19 and 6.20 resolve to a tone within an unambiguous chord region.

MELODIC PITCH HIERARCHIES

The next step is to determine whether tones which are outside the assumed triad are non-chord tones. A particularly conspicuous defect in the pitch-only assessment of chordal compactness is the possibility of misinterpreting what is actually the root as an outlying addition to the triad (or vice versa). The most common confusion involves chords with the apparent members 1-3-5-7 and 1-3-5-6. A 1-3-5-13 chord in the tonic could also be a 1-3-5-7 chord in the key of the submediant if the 13th is transposed down two octaves. Therefore all 6–5 suspensions and many other accented non-harmonic tones are incorrectly identified when only pitch and rhythmic information from the notes within a chord are used. In addition, the optimal rhythmic weighting of root intervals tends to bias this identification incorrectly toward the interval of a sixth as a root. Therefore, the melodic context of a note is needed to help identify the correct root.

All basic forms of non-harmonic tones (e.g., passing tones, neighboring tones, and suspensions) involve resolution into or out of a chord tone. Therefore, if two melodically adjacent tones are found in the same chord region, it is highly likely that one of these is a chord tone and the other is a non-harmonic tone.

It follows that any high-order triadic position (such as a 9th, 11th, or 13th) can be scaled down to be treated instead as a note outside the stack of thirds by attaching it to a lower-order triadic position. For example, previous portions of the root-finding algorithm would identify a chord containing the correct set of intervals 1-3-5-13 as a chord built on the 13th, which generates a more compact set of root intervals: 1-3-5-7. However, if the 13th resolves melodically to the fifth degree of the chord, the 13th can be reinterpreted as the sixth scale degree. Therefore, the correct root configuration of 1-3-5-6 is preferable to the 1-3-5-7 analysis of the sixth scale degree. We call this procedure a pitch-hierarchy assessment (Figure 6.21).
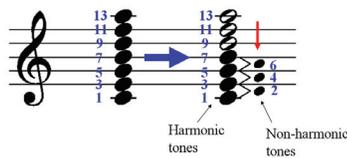
Figure 6.21. *Model for pitch-hierarchy assessment, in which a note with a high triadic chord position is tested to see whether it is connected to a lower position in a different triad, in which case it is more likely to be a non-chord tone than an outlying chord tone.*

We can now add the pitch-hierarchy assessment to the previously defined chord-tone metrics. Figure 6.22 shows revised root calculations for the musical example found in Figure 6.5. Pitches which approach or leave a lower triadic-level note are reassigned lower weights. For example, with the test root on C, the A pitch-class assigned to the 13th in the chord can be reinterpreted as a melodic extension of lower triadic notes, since it precedes the G (in the 5th position) and follows the B (in the 7th position).
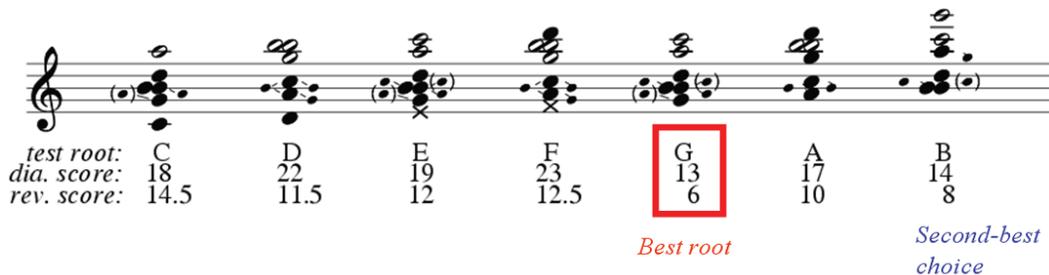
Figure 6.22. *Root-assessment scores revised for pitch hierarchy.*

6.4.3 Key Context

Most roots can be identified by incorporating the previous three contexts into one algorithm. However, there are two cases which can be improved by considering the local key context of the chord. These are the case of a missing root and the occurrence of an ambiguous root. We encounter both in Schubert's posthumously published Theme and Variations on a theme by Anton Hüttenbrenner in A Minor for piano D.576 (Figure 6.23).

Figure 6.23. *The theme (a) and eleventh variation (b) of Schubert's A Minor set, D.576.*

In the variation shown above, the actual root (a) of the downbeat in Bar 1 does not appear until the end of the measure. The ear might hear the downbeat in the context of A Minor because the key has been restated over and over in the preceding variations, but confined to the evidence present, an algorithm will not find it. The phenomenon of a missing root occurs again in the second measure, which follows the pattern of Bar 1. Still more contrary to traditional concepts of tonal harmonic usage is the downbeat of Bar 6 in the variation, where in the absence of the hypothetical root (a) we find a triple instance of the third (c) and a single instance of the fifth (e). In manual analysis of such passages, there would be no absolute consensus among theorists on the root in Bar 5 either, because although it is literally a diminished 7th on the seventh scale degree (g#), it may also be considered a 9th chord on the dominant (E), in which case the missing root on the downbeat of Bar 6 might enable us to "hear" the implied pedal point on E that starts at the beginning of Bar 5 and is supplied by the actual E in the bass on the second half of the first beat of Bar 6.

Among the several existing harmonic-analysis programs we have explored, the highest degree of compatibility with manual analysis seems to exist in Sleator and Temperley's *Melisma Music Analyzer* (2000). This owes partly to its strengths in time segmentation. *Melisma*, the main objective of which is to model music cognition, fixes a

minimum harmonic-rhythm duration (e.g., at the eighth-note level) and incorporates adjacent chord regions in which the notes representing the root remain at constant pitch. The rapid rate of apparent harmonic change in Schubert's theme and variations offers one case in which several divergences from manual analysis occur.

6.5 Analytical Applications

To test the two-angle approach we ran our evaluation procedure on 70 Bach chorales which had been manually analyzed prior to the test.[6] The repertory contained 3,162 chords. To evaluate the accuracy of various components of the root-finding algorithm we ran several tests comparing the outcomes of the manual and automatic procedures. A musical example and its analysis are shown in Figure 6.24.

Figure 6.24. *Automatic chord identification of a four-part Bach chorale with inner voice independence.*

6.5.1 Results

In this series of tests, we found that 39% of the pitch sets constituted simple triads, 26% constituted triads with an added seventh, and 35% included non-harmonic tones. The rearrangement of tones to produce chordal compactness produced an overall accuracy rate of 90%. In the most complex cases, the roots were identified correctly 75% of the time. The number of specific pitch sets encountered totaled 246. Some sample cases and their rate of occurrence are shown in Figure 6.25.

	Case 135	Case 7	Case 2	Case 4	Case 6	Case 24	Case 26	Case 46	Case 246
Occurrence:	40%	25%	12%	8.1%	7.2%	4.2%	0.1%	2.2%	0.4%

Figure 6.25. *Selected pitch sets from among 246 found in 70 Bach chorale harmonizations.*

In conducting these tests we found that while a uniform interval space is elegant because of its lower dimensionality, the lowest error rate with root-interval weights varied independently. If the techniques are optimized for one dataset, they may not work as well with another.

6.5.3 Summary of Error-Rate Calculations

Basic pitch-only information typically gives a root-identification error rate of about 10% for the chorales. By adding rhythm to basic parameterized pitch-space configurations, the error rates improve from about 9% to 8% for duration and to about 6% for metric data. Interval weights randomly optimized to minimize identification errors in the same data reduce the rate to 4.5%. Alternatively, combining duration and metric data yields approximately a 5% error rate, which is close to the maximum possible error rate using optimized interval weights alone. For optimized interval-weight pitch configurations, the error rate improves by only 0.1% when durational data is considered, and from 4.5% to 3.5% when metrical position is taken into account. Adding the melodic-context hierarchy corrections to compensate primarily for accented tones decreases the root-identification error rate to about 1%.

6.6 Future Work

The harmonic analysis of a complete work involves two processes of segmentation. The first consists of identifying a sequence of chord regions. The second involves identifying the root of the chord that dominates each region. Most of our focus is on the second, because it is essential to making the first more efficient. We are still evaluating three possible means of further refinement in this process: (1) the removal of melodically linked tones from the calculation, (2) the reassignment of the weight of a linked tone to the attached chord tone, and (3) the scaling of linked tones in a manner similar to that in which we scale rhythm. There are further needs to balance rhythmic and melodic rules generally, to evaluate low-level rules on other repertories, and to normalize root scores in different time spans.

We would like to experiment with a matching process for regions of a work with tonally similar profiles to establish whether the model of a harmonically well-defined section might serve to disambiguate missing and displaced roots in passages with a thin musical texture. We also hope to explore the issue of particular harmonic rhythms specific to individual genres.

Some experimentation with our procedures has already been attempted for automatic phrasing of electronic data according to Sundberg's *Director Musices* rubrics for expressive performance. (See *http://www.speech.kth.se/music/performance/performance_emotion.html*).

For those who wish to run their own experiments, data in the Humdrum kern format is available at the KernScores website (*http://kern.ccarh.org* and *http://kern.humdrum.net*), which is briefly described in Sapp (2005b). Program source code for the analysis of Humdrum files can be found at *http://museinfo.sapp.org* and at *http://extras.humdrum.org*.

Notes

1. As an example, his Rule 7 reads: "If a sonority is not tertian OR it is accented AND dissonant AND the next sonority is tertian AND the next sonority has a lower tertian dissonance level OR it is unaccented AND dissonant AND the last sonority is tertian AND the last sonority has a lower tertian dissonance level, THEN the sonority is dissonant in context."

2. See *http://www.link.cs.cmu.edu/music-analysis*.

3. The 2007 book was published too recently to permit evaluation of its key-finding approach within this article, but Temperley's own contribution (Ch. 2) pursues some of the topics (particularly "tonalness") introduced in the book.

4. Real-time analysis from Temperley's system can be found for the virtual scores at *http://kern.humdrum.net*.

5. Riemann's most comprehensive work on harmony was published in German in 1893 and translated into English in 1895. His use of the *Tonnetz* in a musical context was borrowed from Ottokar Hostinský's *Die Lehre von den musikalischen Klängen* (1879).

6. Fifty of the chorales were marked as found in music-theory literature as compiled by David Huron. Another 20 were analyzed by me. The virtual scores for the Bach chorales and their harmonic analyses can be downloaded from *http://kern.humdrum.net/cgi-bin/ksbrowse?1=/users/craig/classical/bach/bhchorale*.

References

Aarden, Bret (2003). *Melodic Expectancy*. Ph.D. dissertation, The Ohio State University.

Chew, Elaine (2000). *Towards a Mathematical Model of Tonality*. Ph.D. dissertation, MIT.

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

Euler, Leonhard (1739). *Tentamen novae theoriae musicae* in *Opera Omnia*, ed. F. Rudio, A. Krazer, A. Speiser, and L. G. du Pasquier (Stuttgart: Teubner, 1926), 3/1.

Krumhansl, Carol (1990). *Cognitive Foundations of Musical Pitch*. Oxford, UK: Oxford UniversityPress.

Krumhansl, Carol (2005). "The Geometry of Musical Structure: A Brief Introduction and History," *ACM Computers in Entertainment* 4/3: 1–14.

Lerdahl, Fred (2001). *Tonal Pitch Space*. Cambridge, MA: The MIT Press.

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

Lewin, David (1987). *Generalized Musical Intervals and Transformations*. New Haven, CT: Yale University Press.

Maxwell, John H. (1984). *An Artifical Intelligence Aproach to Computer-Implemented Analysis of Harmony in Tonal Music*. Ph.D. dissertation, Indiana University.

Mickelsen, William C. (1977). *Hugo Riemann's Theory of Harmony*. Lincoln, NE: Univ. of Nebraska Press.

Morris, Robert (1998). "Voice-Leading Spaces," *Music Theory Spectrum* 20/2: 176–208.

Riemann, Hugo (1977). *History of Music Theory*, Book III, tr. William C. Mickelsen. Lincoln, NE: University of Nebraska Press.

Sapp, Craig Stuart (2001). Tonal Landscape Gallery. *http://ccrma.stanford.edu/~craig/keyscape/*

Sapp, Craig Stuart (2005a). "Visual Hierarchical Key Analysis," ACM Computers in Entertainment 3/4. PDF online at *http://ccrma.stanford.edu/~craig/papers/05/p3d-sapp.pdf*.

Sapp, Craig Stuart (2005b). "Online Database of Scores in the Humdrum File Format," Proceedings of ISMIR 2005 (London), pp. 664f. *http://ismir2005.ismir.net/proceedings/index.html*.

Segnini, Rodrigo, and Craig Stuart Sapp (2005). "*ScoreGram*: Displaying Gross Timbre Information from a Score," *Proceedings of the Symposium Series on Computer Music Modeling and Retrieval (CMMR, Pisa, Italy, Sept. 26-28, 2005)*. PDF online at *http://portal.acm.org/citation.cfm?id=1095534.1095544*.

Sleator, Daniel, and David Temperley (2000). *The Melisma Music Analyzer*. *http://www.link.cs.cmu.edu/music-analysis/*

Singer, Mark (2007). "Fantasy for Piano," *The New Yorker,* Sept. 17, 2007, pp. 61–81.

Taube, Heinrich (1999). "Automatic Tonal Analysis: Towards the Implementation of a Music-Theory Workbench," *Computer Music Journal* 23(4): 18–32.

Temperley, David (1996). *The Perception of Harmony and Tonality: An Algorithmic Perspective*. Ph.D. dissertation, Columbia University.

Temperley, David (2001). *The Cognition of Basic Musical Structure*. Cambridge, MA: The MIT Press.

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