A quick guide to terminology and notation

'A' is used as an abbreviation of 'augmented', 'd' for 'diminished', 'M' for 'major', 'm' for 'minor', 'p' for 'perfect'; so 'p5' means 'perfect fifth', 'A4' means 'augmented fourth'. An un-numbered letter of 'U' is used to represent 'unison'.

A forward slash is used to include both an interval **and** its octave inversion, so 'M3/m6' refers to both M3 and m6.

Strict harmonic notation is assumed, so the intervals that make up a major or minor triad are p5/p4, M3/m6 and m3/M6; M2/m7 is taken to represent the difference between p5 and p4, as well as the difference between p4/p5 and m3/M6;[1] m2/M7 is taken to represent the difference between p4/p5 and M3/m6. All other intervals are taken to be augmented or diminished.

An interval is assumed to include all octave expansions unless explicitly stated otherwise; so 'U' includes octave and double octave, 'p5' includes perfect 12th and perfect 19th.

Roman upper case letters are used for major triads and lower case for minor, so 'C' is 'C major triad', 'c' is 'c minor triad'. Within the text, italic letters are used for notes, so '*c*' is 'the note c'. Roman numerals are used for triads, and their number is relative to the parallel diatonic major scale's step number (e.g. in C major, the triad E♭ is notated '♭III'); upper case indicates major, lower case indicates minor. Arabic numerals are used for notes, and their number is relative to the parallel diatonic major scale (e.g. in C major, the note f♯ is notated '♯4').

The superscript suffixes '^{o}' and '$^{+}$' appended to a triad, indicate that the triad is diminished or augmented, respectively.

Directional arrows between pairs of triads or notes indicate whether the relationship goes in either direction, or just the one indicated; so C → D indicates a progression from C to D, while C ↔ D indicates a progression from either C to D, or D to C.

Notes are illustrated either on a Wicki/Hayden-style isomorphic keyboard, or on an isomorphic lattice. In the lattice, octave equivalence is assumed, the near-vertical lines connect p5/p4, the near-horizontal solid lines connect M3/m6, the near horizontal dotted lines connect m3/M6. The lattice should be understood to be a 2-dimensional projection of a toroidal lattice whose syntonic comma and enharmonic diesis (A7/d2) are mapped to unison. To recreate the toroid, the flat lattice needs to be looped so that notes which differ by the syntonic comma and enharmonic diesis lie on top of each other. For this reason, in the 'unwrapped' 2-D lattice, the **same** note may be shown in **two** locations.

Tonal Function in Harmonic Scales

Andrew Milne

When two notes are played successively, we have a *melodic* interval; when two notes are played simultaneously, we have a *harmonic* interval. Our cognition of melodic and harmonic intervals is subtly different, and each provides its own set of expectations and restrictions that act as cognitive 'forces'. It is from the interaction of these forces that tonal function emerges.

[1] These two types of major second differ by an interval called the *syntonic comma*. In this chapter, the syntonic comma is ignored because it has no cognitive significance.

Melodic prototypes

When two notes are played in succession, and their frequency is identical, they are heard to have the same pitch. If there is a very small frequency-difference between the two notes, they will still be heard as identical;[2] as the frequency-difference increases the listener will begin to hear the second note as an *alteration* of the first note; as the frequency-difference increases still further, the second tone will begin to be heard as unrelated to the first note, and will now sound like a *new* and unrelated tone rather than as an alteration of the first.[3] The boundary between alteration and newness occurs between the semitone and tone; the melodic semitone is heard as an alteration, the whole-tone is not.

The unison can be considered to be a *prototypical* interval, because it is easily recognisable and any semitone-like deviation from it will be heard as an alteration of it. Because musical (pitched) tones are made up of harmonic partials, intervals with simple ratios share more partials than intervals with complex ratios;[4] they are made up of more of the same 'stuff'. This means that semitone-like deviations from simple ratios are also subject to being heard as alterations, and so these simple ratios are also prototypical.[5]

In any conventional tuning system that has no step size smaller than a semitone,[6] and **devoid of any *harmonic context***,[7] the intervals of: unison (which approximate frequency ratios of $1:2^{(n-1)}$); perfect fifth (which approximate frequency ratios of $2:3\times2^{(n-1)}$); and perfect fourth (which approximate frequency ratios of $3:2^{(n+1)}$), are heard as *prototypical*. They are prototypical because any intervals that differ by a semitone from them are heard as *alterations* of them rather than as independent entities in-themselves; so semitone-like intervals are heard as alterations of unison and tritone-like intervals are heard as alterations (by a semitone) of p5/p4.

Melodic intervals similar in size to the major third and the minor sixth approximate simple frequency ratios ($4:5\times2^{(n-1)}$ and $5:2^{(n+2)}$, respectively), but they are also only a semitone from the prototypical p4 and p5 ratios. For this reason, these intervals are inherently ambiguous because they can be heard in two possible ways – as prototypes or as alterations. We shall see later how it is the harmonic context which defines which way any given M3/m6-like interval is actually heard. The effect of harmonic context on melodic intervals is explained fully in 'Normalisation of alterations', on page 6.

[2] The point at which an observer correctly identifies that a frequency has changed on at least 76% of occasions is known as the *frequency difference limen*, which according to Rossing (1990) is approximately 8 cents, though in the far-from-ideal conditions under which music is listened to it is likely to be somewhat higher.

[3] When a sound is played at an audible frequency, it causes hairs in the cochlea to vibrate. For any given frequency, the hairs over a small area will vibrate; so two frequencies that are close will cause a number of the same hairs to vibrate. This 'made of the same stuff' quality unifies their identity.

[4] For any interval of ratio p:q, and assuming harmonic partials, the ratio of *partials common to both notes* to *all partials* is $2/(p+q)$.

[5] For a more complex model of this, see Milne (2005).

[6] This includes 12-tone equal temperament and all meantone tunings.

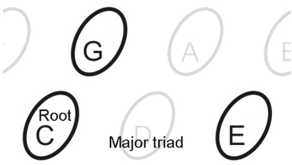
[7] Harmonic context does not have to be explicit because it can be implicitly imagined by the listener.

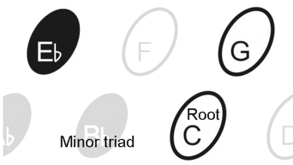
Alterations are heard as “mistuned” prototypes, but this mistuning can be *justified* by *active resolution*. In active resolution, the altered note continues in the direction of its alteration to the next available note that is not in an altered relationship with the first; in this way the alteration is justified as a melodic “stepping stone” between two unaltered notes.

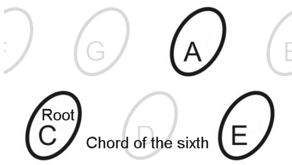
Harmonic prototypes

When played simultaneously, certain sets of notes sit comfortably together, they are stable and they are heard to be in some way *unified* in that they form an entity that is more than just the sum of its parts. These unified and consonant entities are commonly recognised in music theory as the three-note *major* and *minor* triads. The major triad is generally recognised to be stable only in its root position, and the minor triad in either its root position or its first inversion (the first inversion minor triad is called a *chord of the sixth*).

These three types of triad are prototypical (or *proto-triads*), because any other type of triad (e.g. second inversions, augmented, diminished, suspended triads, etc.) is heard as an alteration of one of them. In a proto-triad, each note’s pitch is *supported* by the other two notes. The three proto-triads, played in close position on an isomorphic keyboard, are shown below:



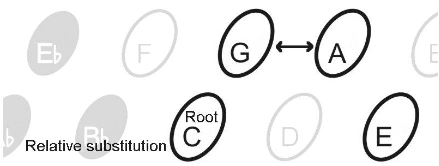


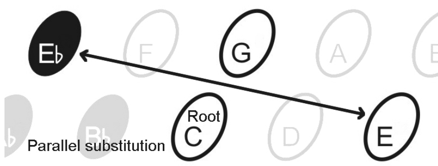


The root (which is the bass note of the three proto-triads) is the principal *pitch-identifier* of the triad. The pitch-identifier is the note that best represents the pitch of the triad as a whole; the note that a casual listener would sing to imitate the chord. A proto-triad is heard as if it has a single pitch, which corresponds to its root, but with added ‘major’ and/or ‘minor’ character.

Same-root substitution

Any two proto-triads that share the same root can be readily *substituted*, one for the other. A major triad can be readily substituted with a chord of the sixth with the same bass note (and vice versa); this is called *relative* substitution. A major triad can be readily substituted for a minor triad with the same root (and vice versa); this is called a *parallel* substitution.[8] A minor triad can also be substituted with a chord of the sixth (and vice versa), although this involves two moving tones rather than one, and so is less suitable as a substitution.





[8] Confusingly, the German tradition, using terminology derived from Hugo Riemann, uses the term ‘parallel’ where the English tradition (as used in this chapter) uses ‘relative’ (Gjerdingen in Dahlhaus, 1990, p. xii).

Because they all share a common root (pitch-identifier), these substitutions are like different ways of saying the same thing. This means that one can easily substitute one *same-root* proto-triad for another.

Alteration in a harmonic context

For a note in a triad to be heard as an alteration, it must be possible to *extrapolate*, with the mind, another triad that contains the note from which this alteration has been made. For this pre-altered triad to be easy to extrapolate it must be a proto-triad and, because alteration has a range of no more than a semitone, the pre-altered note(s) must be no more than a semitone from the altered note(s). This eliminates the relative substitution as a means to effect a harmonic alteration (because it moves by a whole-tone), but it allows for the all-important parallel substitution.

There are other pairs of triads that can serve as alterations, because their elements differ by a semitone. None of them, however, are same-root substitutions so they cannot be imposed with the same seamless elegance as the parallel substitution. Including the parallel, there are three possible types of harmonic alteration, these are shown below.

> For any triad there is one other proto-triad whose third differs by a semitone:
>
> C ↔ c, the same-root *parallel* alteration.
>
> For any triad there is one other proto-triad whose root **and** fifth differ by a semitone:
>
> C ↔ c♯, the semitone-root *Neapolitan* alteration.
>
> For any triad there is one other proto-triad whose root **and** third **and** fifth differ by a semitone:
>
> C ↔ C♯, c ↔ c♯, the semitone-root *shift* alteration.

For any triad there are **no** other proto-triads whose root, **or** fifth, **or** root **and** third, **or** third **and** fifth differ by a semitone.

This gives us the three types of possible triadic alterations: parallel (C ↔ c), Neapolitan (C ↔ c♯), and semitone shift (C ↔ C♯ and c ↔ c♯).

Parallel, Neapolitan and shift substitutions are the harmonic means of articulating an alteration. Of these three substitutions, the parallel is by far the most readily extrapolated because it is a same-root substitution, and so is heard to be in concordance with the composer/performer's intention.

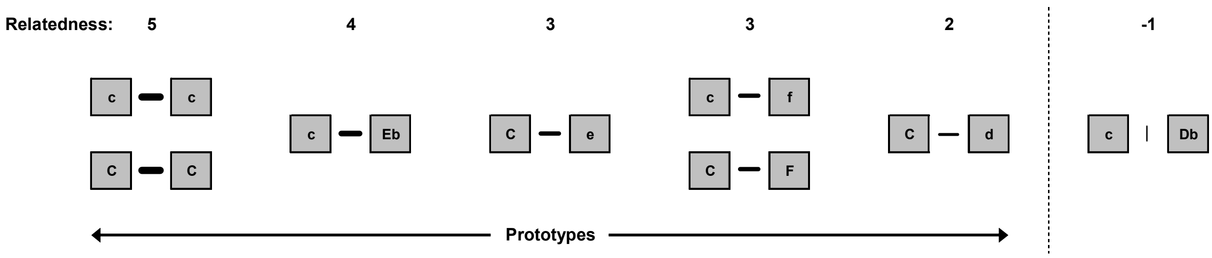
Triadic relatedness

Prototypical progressions

Any pairing of two triads has a total of nine melodic relations that occur between the notes of one triad and the notes of the other triad: the root of the first triad with the root, third and fifth of the second triad; the third of the first triad with the root, third and fifth of the second triad; the fifth of the first triad with the root, third and fifth of the second triad. The overall *relatedness* of two triads can be estimated by comparing the number of prototypical (U, p5/p4) melodic relations, to the number of alterations of those prototypes (semitones and tritones).[9] Pairs with a positive relatedness can be considered to be inherently well-related, and those with a negative relatedness to be inherently poorly-related. There are twenty-six possible triad pairings, and the following chart ranks them in order of relatedness:[10]

<table>
<tr><th colspan="3">Progression</th><th>Relatedness</th></tr>
<tr><td></td><td>C ↔ C</td><td></td><td>5</td></tr>
<tr><td></td><td>c ↔ c</td><td></td><td>5</td></tr>
<tr><td>vi | ii | iii | iv</td><td>c ↔ E♭</td><td>I | IV | V | ♭VI</td><td>4</td></tr>
<tr><td>I | IV | ♭VI</td><td>C ↔ e</td><td>iii | vi | i</td><td>3</td></tr>
<tr><td></td><td>C ↔ c</td><td></td><td>3</td></tr>
<tr><td>V | I</td><td>C ↔ F</td><td>I | IV</td><td>3</td></tr>
<tr><td>i | v</td><td>c ↔ f</td><td>iv | i</td><td>3</td></tr>
<tr><td>ii</td><td>c ↔ F</td><td>V</td><td>2</td></tr>
<tr><td>V | I</td><td>C ↔ d</td><td>vi | ii</td><td>2</td></tr>
<tr><td>V</td><td>C ↔ f</td><td>i</td><td>1</td></tr>
<tr><td></td><td>C ↔ E</td><td></td><td>0</td></tr>
<tr><td></td><td>c ↔ e</td><td></td><td>0</td></tr>
<tr><td></td><td>C ↔ E♭</td><td></td><td>0</td></tr>
<tr><td></td><td>c ↔ e♭</td><td></td><td>0</td></tr>
<tr><td>IV</td><td>C ↔ D</td><td>V</td><td>0</td></tr>
<tr><td>ii</td><td>c ↔ d</td><td>iii</td><td>0</td></tr>
<tr><td>iii</td><td>c ↔ D♭</td><td>IV</td><td>-1</td></tr>
<tr><td></td><td>C ↔ c♯/d♭</td><td></td><td>-2</td></tr>
<tr><td></td><td>c ↔ F♭/E</td><td></td><td>-2</td></tr>
<tr><td>V</td><td>C ↔ D♭/C♯</td><td>♭VI</td><td>-3</td></tr>
<tr><td></td><td>c ↔ d♭/c♯</td><td></td><td>-3</td></tr>
<tr><td></td><td>C ↔ f♯</td><td></td><td>-3</td></tr>
<tr><td>iv</td><td>c ↔ D</td><td>V</td><td>-3</td></tr>
<tr><td></td><td>C ↔ e♭</td><td></td><td>-4</td></tr>
<tr><td></td><td>C ↔ G♭/F♯</td><td></td><td>-5</td></tr>
<tr><td></td><td>c ↔ g♭/f♯</td><td></td><td>-5</td></tr>
</table>

If a parallel, Neapolitan, or shift alteration is applied to either, or both, triad(s) in these pairings, the relatedness of the new altered pair is either increased or decreased. The following eight triad pairs are those whose relatedness cannot be improved by any of the three types of alteration, and they are ranked from left to right according to their relatedness:



The first seven of these pairs have positive relatedness; only the eighth pair (c ↔ D♭) has negative relatedness. These first seven triad pairs represent *prototypical progressions* because

[9] Because M3/m6-like intervals can be interpreted as being either prototypical or as alterations of p5/p4 (see on page 2) their contribution to relatedness is inherently ambiguous, so they cannot be used as part of its calculation.

[10] There are 26 possible triad pairs as long as enharmonically ambiguous pairs are considered to be identical (e.g. C ↔ D♭ ≡ C ↔ C♯); if they are not, there are 32 possible triad pairs.

they cannot be improved by alteration **and** they are inherently well-related. Therefore, all non-prototypical triad progressions (with the exception of the 'difficult' eighth progression c ↔ D♭, or iii ↔ IV) are heard as alterations of these seven prototypes.[11]

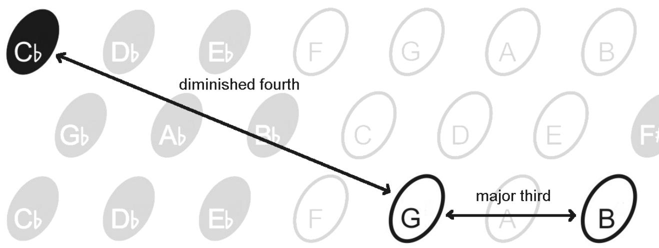
Normalisation of alterations

All intervals that are invoked by these prototypical progressions are considered to be *normalised* – i.e. any potential alterations are no longer heard as such, because, within their immediate triadic context, they are **supported** (they are members of a triad) and **constrained** (they cannot be altered without damaging relatedness), and the triads in which the interval is expressed are **mutually supportive** (they are well-related).

The interval m2/M7 occurs in four of the prototypical progressions; the intervals AU/dU, A4/d5, and A5/d4 occur in none of the prototypical progressions. This means that m2/M7 is a normalised interval, because it is unalterable in well-related triads; but AU/dU, A4/d5, and A5/d4 are un-normalised, because they are alterable in well-related triads, and unalterable only in poorly-related triads. For instance, in a prototypical progression like C ↔ F, there is no alteration of either C or F that can turn the m2/M7 between the third of C and the root of F into a unison; but in a non-prototypical progression like C ↔ E the AU/dU between the fifth of C and the third of E can be made into a unison by altering E to e; in the progression c ↔ D♭ the A4/d5 between the fifth of c and the root of D♭ cannot be turned into p5/p4 by alteration, but at the same time the two triads are poorly related so insufficient support is given to this interval to normalise it.

All augmented and diminished intervals are ultimately heard as either direct or indirect alterations of U or p5/p4. AU/dU is heard as a direct alteration of U (m2/M7 is not); A5/d4 is heard as a direct alteration of p5/p4 (M3/m6 is not); A4/d5 is heard as a direct alteration of p5/p4; the remaining intervals, A2/d7, A6/d3, A3/d6 and A7/d2,[12] *imply* an alteration of p5/p4 or U, because they cannot be invoked without also invoking one of the former three aug/dim intervals.

Conventional notation, therefore, does us the great favour of discriminating between aug/dim intervals and their enharmonic equivalents because, even though they are represented by the same tuning in most conventional temperaments, their effect is quite different.[13] Conventional keyboard and fretboard instruments, however, do not discriminate between enharmonic equivalents, but on an isomorphic keyboard the difference is quite explicit:

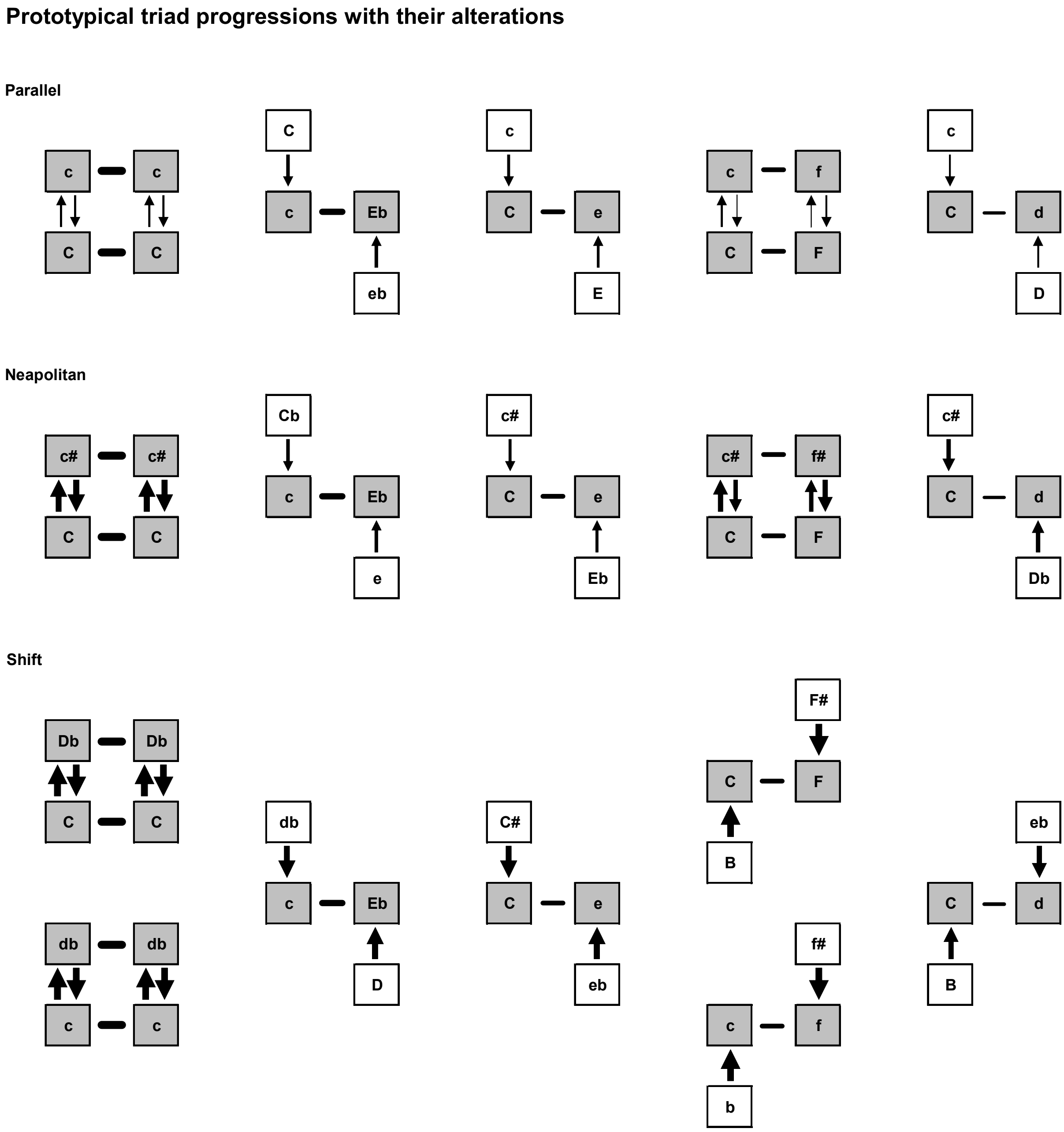


[11] For more information about this progression, see page 2.

[12] The following intervals can be invoked by two triads: U; p5/p4; M2/m7; m3/M6; M3/m6; m2/M7; A4/d5; AU/dU; A5/d4; A2/d7; A6/d3; *A3/d6*; *A7/d2*. The latter two italicised intervals can only be invoked when unusual, though possible, enharmonic notation is chosen.

[13] The difference in effect has nothing to do with the tuning size of the interval, and everything to do with its harmonic context.

The following chart shows how every possible proto-triad progression is interpreted; the prototypical progressions are shown connected with a horizontal line (whose thickness indicates their degree of relatedness); the alterations are shown above and/or below and are connected to their prototypes with a directional arrow (whose thickness indicates the degree of improvement in relatedness that the substitution gives); where there is a choice of possible alterations, preference is given firstly to parallel, secondly to Neapolitan, thirdly to shift.



For example, the chart shows: C → c is heard as an alteration of C → C (line 1: column 1); E♭ → C is heard as an alteration of E♭ → c (line 1: column 2); C → E♭ is heard as an alteration of C → e (line 2: column 3); c → E is heard as an alteration of C → e (line 1: column 3); C → e♭ can be heard either as an alteration of c → E♭ (line 1: column 2) or as an alteration of C → d (line 4: column 5); C → D is heard as an alteration of C → d (line 1: column 5); C → F♯ is heard as an alteration of C → F or C → G (line 3: column 4); etc..

Tonality

Tonal-harmonic cadences

A cadence is the means by which music is brought to a point of ‘rest’, ‘closure’, ‘completion’ or ‘resolution’. A cadence can be articulated using purely rhythmic or melodic devices, but it can also be articulated, perhaps most powerfully of all, with harmony. The *harmonic cadence* is the most ‘distilled’ expression of harmonic tonality; it is a short chord progression that expresses disturbance followed by resolution or, as Lowinsky puts it, “the cadence is the cradle of tonality”.[14]

A harmonic cadence is effected by invoking an aug/dim interval, which is then actively resolved. This can be done with a progression of three triads: *antepenult → penult → tonic*, in which case the aug/dim interval is invoked in the transition from antepenult to penult, and actively resolved in the transition from penult to tonic. It can also be done with two chords: *penult → tonic*, in which case the aug/dim interval is invoked as a harmonic dissonance within the penult,[15] and this is resolved in the transition to the tonic.

The penult is an alteration (preferably parallel) of a triad that is in a prototypical relationship with the antepenult. When this is the case, the penult is **easy** to hear as a **resolvable** alteration. The penult resolves by continuing the motion set up by the alteration to the next available scale note.

The ubiquitous IV → V → I cadence is a classic example: the IV → V progression is heard as a parallel alteration of IV → v, and the altered third of V (scale note 7) resolves in the direction of its alteration to scale note 1 (the root of I). This also explains why reversing the ‘subdominant’ and ‘dominant’ triads to V → IV → I does not provide a satisfactory cadence, because IV has no alteration that is in a prototypical relationship with V. The directional nature of parallel and Neapolitan alterations determines the directional nature of tonal cadences. Similarly, this also explains why using the dominant’s relative in IV → iii → I does not work, because iii has no alteration that is in a prototypical relationship with IV.

All of the conventional cadences, such as ii/IV → V → I/i, iv → V → I/i, ii/IV → V → vi, etc. are explained by parallel alteration. In each of these cadences, the V triad is heard as an alteration of v. Unusual cadences such as ♭III → ♭II → I (flamenco cadence) and iii → ♭II → I can be explained by Neapolitan or shift alterations: in the former, the penult is heard as an alteration of ii; in the latter, the penult is heard as an alteration of II.

When an aug/dim interval is invoked between two triads, for one of the triads to function as an effective penult, the note that invokes the tritone must be that triad’s third (or root **and** fifth, or root **and** fifth **and** third). When a tritone interval is invoked by just the root **or** just the fifth, the prototype cannot be easily extrapolated so it loses much of its normal activity and its effect is clumsy without special care. Note, for example, the special rules surrounding the use of the iii triad in traditional music theory; the following quote and its location exemplify this, in that this apparently simple diatonic triad is not covered until Grade VIII

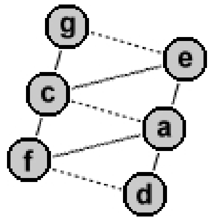
[14] Lowinsky, 1961, p. 4.

[15] That is, a chord that contains an aug/dim interval, such as dominant seventh, augmented sixth, augmented or diminished triad, etc.

(final) of The Associated Board of the Royal Schools of Music syllabus: "The triad on the mediant (III) is a chord which needs careful use.…The leading-note of the scale seems to lose something of its character when it becomes the fifth of the mediant chord. It *falls* quite naturally in the progressions III to VI and III to IV".[16]

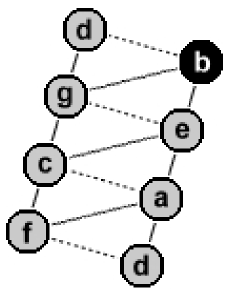
Tonal analysis of scales

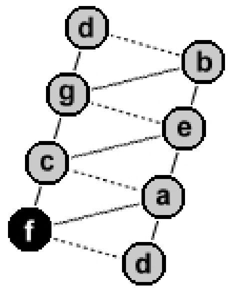
Up to this point, a lot of information has been generated, so it needs to be condensed into a more digestible form. We also need to have a means of analysing not just three-triad cadences but the gamut of triadic relationships that are found in any given scale. The means of doing this is to construct the *maximal* scale that contains **only prototypical progressions**.[17] In such a scale, all available triad progressions are stable and well-related, and it represents the greatest set of notes for which there is **no tonal function**. This scale is the *diatonic hexachord*, and it is illustrated below: [18]



Because the diatonic hexachord contains only prototypical triad progressions, it can be regarded as the prototypical tonal harmonic scale, in relation to which any *extraneous* tones are heard as alterations. This conception of the diatonic hexachord gives us a method to analyse the tonal function of any given scale.

To analyse any given scale using the *hexachord method* requires finding all available *underlying* hexachords. An underlying hexachord is one that contains the maximum possible number of the scale's notes, and for all seven-note scales, there will be more than one underlying hexachord. For instance, the diatonic scale has two underlying hexachords, as indicated with the grey notes:





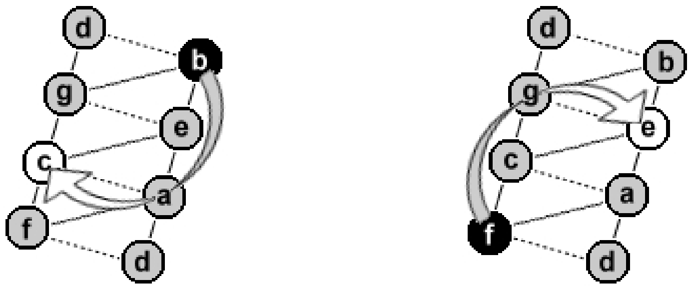
[16] The Associated Board of the Royal Schools of Music, 1958, p. 109-110.

[17] The term 'maximal' refers to the scale with the largest number of notes that still fulfils the imposed condition.

[18] The diatonic hexachord is structurally identical to the "Guidonian hexachord" which was used in medieval music theory and practice, and was the origin of the solfege system.

[20] Indeed, this superfluousness explains the weakness of musical lines that proceed from a note to its alteration to its resolution, rather than substituting the first note with its alteration. For example, compare G7 → C → F → f → C to G7 → C → f → C.

Each underlying hexachord will **not** include one, or more, of the scale's notes; these notes are extraneous to this hexachord and so are heard as *active* alterations. The notes to which these altered notes actively resolve are *resolution* notes. This gives the *functional status* of the scale's notes for each underlying hexachord. In the chart below, active notes are black, resolution notes are white, and all other notes are grey:



Functional consistency

To check for *functional consistency* we need to check that the status (active or resolution) of each note in every available underlying hexachord is consistent. To do this, we create a *composite hexachord* where each note is indicated so that: any note that is extraneous in any hexachord is an active note in the composite hexachord; any note that is a resolution note in any hexachord, but is an active note in no hexachord, is a resolution note in the composite hexachord. For a cadence following an antepenult → penult → tonic formula, the composite hexachord allows us to analyse tonal function in the following way:

> for a triad to function as a **tonic** it must contain as many of the resolution notes, and as few of the active notes, as possible;
>
> > the **tonic** function of such a triad is stronger if a resolution note is the root (the pitch-identifier of the triad), though it is possible for tonic function, of a less emphatic kind, to be conferred when a resolution note is not the root;
>
> for a triad to function as a **penult** it must contain the active note that leads to the resolution note;
>
> > the active note in the **penult** will be preferably that triad's third, or the active notes will be that triad's root **and** fifth, or they will be all three of the triad's notes (this is to ensure that the triad can be heard as a justifiable alteration);
>
> for a triad to function as an **antepenult** it must contain the active note that is in an aug/dim relationship with the penult's active note;

Positive tonal feedback

Once the tonic is established, it can exercise *positive feedback* to support its status, in the following way: the tonality will inhibit the activity of any active note whose prototype is in an aug/dim relationship with the tonic itself, because such prototypes will be harder to extrapolate. So in the diatonic major scale, for example, choosing 1 as the tonic will actually make 4 less active than 7, while choosing 6 as the tonic will have less effect on whether 4 or 7 are more active. Assuming the chosen tonic triad is validated by the composite hexachord, any hexachord that has an active note that is the same as, or p5/p4 from, the tonic's root, becomes a less significant cognitive foundation.

Tonal scales

So, which scales to analyse? Ideally, we want to analyse scales that are ‘coherent’ and ‘complete’ in themselves. These qualities emerge from a scale providing *sufficient* but not *redundant* or *superfluous* resources for the functions it must fulfil. Those functions are melodic, harmonic, and tonal.

For a scale to serve as a melodic resource, it needs to contain notes that are far enough apart to be easily recognised as different (i.e. the smallest step size needs to be towards the 100% recognisable end of the frequency difference limen).

For a scale to function as a harmonic resource, all its members must be part of at least one major or minor triad. Any non-triadic scale note is superfluous.

For a scale to be used as a tonal resource it cannot contain both a triad and its alteration, because this means that one of the triads will not be heard as a new element, but as an alteration of a pre-existing one. One of the two triads, therefore, will be heard as redundant.[20] This can be simplified by saying that the scale cannot contain a chromatic semitone (AU/dU).[21]

There are five maximal scales that fulfil these conditions: the *prime scales*.

[21] A few useful scales break some of these rules, and it must be understood that the notion of prime scales is just being used to provide a limited subset of useful scales to analyse, not as a blanket prescription.

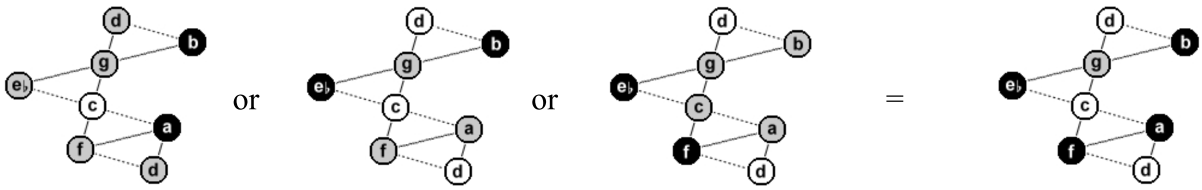
The following diagram provides a hexachordal analysis of the prime scales (the white notes are resolution notes, the black notes are active, the grey notes are neutral):

Diatonic

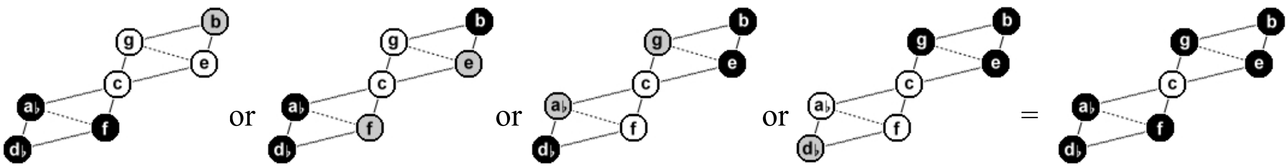
Harmonic minor

Harmonic major

Melodic



Double harmonic



Diatonic: the two resolution notes are *c* and *e*, so there are two potential tonics – C and a. Of these two, C is the stronger because the root is a resolution note (in a, the root is not a resolution note).

> **C diatonic major**: the best available penult is G, and the best available antepenults are F or d.
>
> **a diatonic minor**: the best available penults are d and G, so the best available antepenults of d are e or G, and the best available antepenults of G are F and d.

c harmonic minor: the three resolution notes spell out c as the only available tonic triad. The best available penult is G, and the best available antepenults are f followed by A♭.

C harmonic major: the three resolution notes spell out C as the only available tonic. The best available penult is G, and the best available antepenult is f.

Melodic: the two resolution notes cannot co-exist in the same triad, so the best available tonics will also contain as few active notes as possible; this gives c and G. Of these two c is the stronger because its root is a resolution note (in G, the resolution note is the fifth).

> **c melodic minor**: the best available penult is G, and the best available antepenults are F or d.
>
> **G melodic major**: the best available penult is c, and the best available antepenults are d or F.

Double harmonic: there is only one resolution note, and there are just two possible tonic triads – C and f. Of these two, C is the stronger because its root is the resolution note (in f, the resolution note is the fifth).

> **C double harmonic major**: the best available penult is D♭,[22] and the best available antepenult is e.
>
> **f double harmonic minor**: the best available penult is e,[23] and the best available antepenult is D♭.

The hexachordal analysis shows that, from the five prime scales, there are eight 'modes' that have a tonic-function I or i triad, and so can be considered to be tonal-harmonic scales.

This type of analysis can be simplified even further with the following rule-of-thumb, which can be simply applied to the majority of scales: in any given scale, find the tritone(s) and the major or minor third(s) to which the tritone(s) resolve; the tonic(s) will be the triad(s) that contain one or more of these resolved thirds; the penult will be the triad which contains a tritone note (with the exception of tritone notes which are the same as, or a perfect fifth from, the root of the tonic triad); the antepenult will be the triad(s) which contains the note which is a tritone from the penult's third. This simple formulation provides a quick means of deducing

---

[22] This is one of the rare occasions where the prototype is derived from a shift substitution, so that D♭ is heard as an alteration of D.

[23] This is another shift substitution, so that e is heard as an alteration of e♭

tonal function for unfamiliar scales. However, only the hexachordal method can give a full insight into the scale's tonality.

The four major-tonic scales are diatonic major, harmonic major, melodic major and double harmonic major. Combining these four scales gives the following notes:

1, ♭2, 2, 3, 4, 5, ♭6, 6, ♭7, 7 (Do, Ra, Re, Mi, Fa, So, Le, La, Te, Ti)

The four minor-tonic scales are diatonic minor, harmonic minor, melodic minor and double harmonic minor. Combining these four scales gives the following notes:

1, 2, ♭3, 4, ♯4, 5, ♭6, 6, ♭7, 7 (Do, Re, Me, Fa, Fi, So, Le, La, Te, Ti)

Combining all eight scales gives:

1, ♭2, 2, ♭3, 3, 4, ♯4, 5, ♭6, 6, ♭7, 7 (Do, Ra, Re, Me, Mi, Fa, Fi, So, Le, La, Te, Ti)

This is the standard *harmonic chromatic scale*[24] that provides the resource for the most frequently used *primary*[25] chromatic harmony in common practice music.

Of the eight scales discussed above, only four are 'common currency': the diatonic major scale, and the three minor scales (natural, harmonic and melodic). The remaining four scales are virgin territory, inviting exploration for the intrepid musical explorer. In the next section, we will explore one of these unusual scales in more depth, and see what insights the above method of analysis affords.

The melodic major scale

The melodic major scale has the following notes and triads:

Notes: 1, 2, 3, 4, 5, ♭6, ♭7 (Do, Re, Mi, Fa, So, Le, Te)

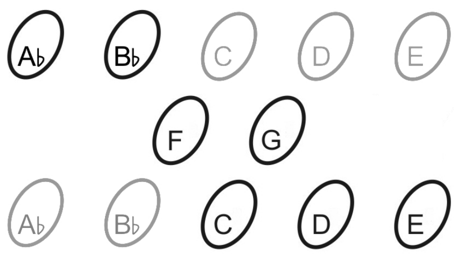
Chords: I, ii^o, iii^o iv, v, ♭VI+, ♭VII

So c melodic major is:

Notes: *c*, *d*, *e*, *f*, *g*, *a*♭, *b*♭

Chords: C, d^o, e^o, f, g, A♭$^+$, B♭

On an isomorphic keyboard, the scale looks like this:



[24] Macpherson, 1920, p. 271.

[25] As opposed to *secondary* chromaticisms, which are introduced by cadential progressions to non-I/i triads.

This is an interesting scale because in many ways it contradicts what we expect of a tonal scale when we are accustomed to just the standard major and minor keys. For instance, this scale has a minor dominant, which is unusual in a tonal context: “without the leading tone, the dominant effect is drastically weakened …. The minor dominant commonly occurs as companion to the descending melodic minor scale, but only exceptionally does the tonic triad follow it.”[26] This is precisely how the dominant functions in the melodic major scale; it proceeds downwards to the minor subdominant and acts as an antepenult rather than as a penult. In this, and many other ways, the melodic major scale functions in an opposite, or complementary, fashion to the diatonic major scale:

In the diatonic major scale, the dominant functions as the penult and the subdominant functions as an antepenult; in the melodic major scale, the dominant functions as an antepenult, while the subdominant functions as the penult.

In the diatonic major scale, we have the cadential progressions ii/IV → V → I, which are heard as an alteration of ii/IV → v → I; in the melodic major scale, we have the cadential progressions ♭VII/v → iv → I, which are heard as an alteration of ♭VII/v → IV → I.

In the diatonic major scale, the penult is heard as an upwards parallel alteration (minor to major); in the melodic major scale, the alteration is heard as a downwards parallel alteration (major to minor).

In the diatonic major scale, the active note (the major penult’s third) resolves upwards to the tonic triad’s root; in the melodic major scale, the active note (the minor penult’s third) resolves downwards to the fifth of the tonic triad. In terms of mood and character, the melodic major’s downward resolution gives a ‘mournful’, ‘yearning’, ‘falling’ quality to the cadence; and, with the resolution ending on the tonic’s fifth rather than root, the cadence is ‘soft’ in character unlike the ‘vigorous’ and ‘definitive’ diatonic major cadence.

In the diatonic scale, there is a major tonic and a minor tonic and it is relatively easy to swap tonality from one to the other (i.e. swapping between the major scale and the relative Aeolian mode a minor third below), though the major tonic is the stronger; in the melodic scale there is a major tonic and a minor tonic that can also be easily swapped (melodic minor and its relative melodic major a perfect fourth below), but this time the minor tonic is the stronger. When using the melodic major scale, one must guard against inadvertent modulation to the stronger minor tonic, just as in the diatonic scale one must guard against inadvertent modulation from the Aeolian mode to the stronger relative major.

Another interesting feature of this scale is its similarity to the diatonic Mixolydian mode, from which it differs by just one note: the flattened sixth. The only diatonic modes that can function as tonal-harmonic scales are the diatonic major (Ionian) and diatonic minor (Aeolian), and any attempt to improve tonal function by altering the *characteristic notes* of the diatonic modes (the two notes that form the diatonic tritone, Fa and Ti), will change the mode;[27] indeed, it was this very process whereby musica ficta changed the medieval modes into our two tonal modes: “The written image of polyphonic music of the Middle Ages and

---

[26] Piston, 1994, p. 53-54.

[27] The characteristic notes are the two notes that define the mode so, by definition, they cannot be altered without the mode changing.

Renaissance conceals the inroads of a nascent tonality upon the modal structure. For these inroads were made largely in performance through observation of the rules of musica ficta."[28] So, to improve tonal function we need to look at alterations that do not touch the diatonic tritone. We can consider, therefore, the melodic major scale to be a 'tonally enhanced' form of the Mixolydian mode – Mixolydian ♭6. Similarly, we can consider the harmonic minor scale to be a tonally enhanced Aeolian mode – Aeolian ♮7, and the melodic minor scale to be a tonally enhanced Dorian mode – Dorian ♮7.

Conclusion

We hear melodic intervals, harmonic structures, and the relationships between harmonic structures, in relation to cognitive prototypes. Deviations from these prototypes create expectations of resolution, and this is the birth of tonal 'force'. Tonal function, as described by traditional music theory, emerges from the complex interaction of these tonal forces. An understanding of this process allows for the conventional rules of tonal function to be deduced from first principles, rather than having to be learned by rote. Furthermore, this hypothesis allows us to describe and explain tonalities that have been largely ignored by traditional theory, such as the melodic major scale and the flamenco cadence. And, although beyond the scope of this chapter, it also provides explanations and descriptions of the tonal relationships encountered when modulating between different tonal centres, as well as opening up a theoretical gateway to possible new forms of tonality based on harmonies with higher-limit consonances.

Bibliography

Dahlhaus, Carl. *Studies on the Origin of Harmonic Tonality*. Trans. Robert O. Gjerdingen (Oxford: Princeton University Press, 1990)..3

Lowinsky, Edward E. *Tonality and Atonality in 16th Century Music*. (Berkeley and Los Angeles: University of California Press, 1961) 8, 15

Macpherson, Stewart. *Melody and Harmony*. (London: Joseph Williams, 1920)..13

Milne, Andrew. 'A model for the cognition of melodic intervals.' *The Tonal Centre*, 2005. <http://www.andymilne.dial.pipex.com/xxx.htm> 2

Piston, Walter. *Harmony*. Revised and expanded by Mark DeVoto. (London: Victor Gollancz, 1994).............14

Rossing, Thomas D. *The Science of Sound* 2nd Ed. (Addison-Wesley, 1990)..2

The Associated Board of the Royal Schools of Music. *Rudiments and Theory of Music*. (England: The Associated Board of the Royal Schools of Music, 1958). 9

[28] Lowinsky, 1961, p. 4.