# The Institute For Research In Cognitive Science

## The Well-tempered Computer

by

**Mark Steedman**

**University of Pennsylvania**
**3401 Walnut Street, Suite 400C**
**Philadelphia, PA 19104-6228**

**November 1994**

Site of the NSF Science and Technology Center for Research in Cognitive Science

**IRCS Report 94-20**

*University of Pennsylvania*
Founded by Benjamin Franklin in 1740

# The Well-tempered Computer

By Mark Steedman

*Department of Computer and Information Science, University of Pennsylvania, Philadelphia PA 19104-6389, U.S.A.*

The psychological mechanism by which even musically untutored people can comprehend novel melodies resembles that by which they comprehend sentences of their native language. The paper identifies a syntax, a semantics, and a domain or "model". These elements are examined in application to the task of harmonic comprehension and analysis of unaccompanied melody, and a computational theory is argued for.

To appear in *Philosophical Transactions of the Royal Society, Series A.*
Key terms: music; computational analysis; key-analysis; harmony; consonance; grammar of music.

# The Well-tempered Computer

By Mark Steedman

*Department of Computer and Information Science, University of Pennsylvania, Philadelphia PA 19104-6389, U.S.A.*

The psychological mechanism by which even musically untutored people can comprehend novel melodies resembles that by which they comprehend sentences of their native language. The paper identifies a syntax, a semantics, and a domain or "model". These elements are examined in application to the task of harmonic comprehension and analysis of unaccompanied melody, and a computational theory is argued for.

---

## 1. Introduction

The question of what constitutes musical experience and understanding is a very ancient one, like many important questions about the mind. The answers that have been offered over the years since the question was first posed have depended on the notion of mechanism that has been available as a metaphor for the mind.

For Aristotle, and for the Pythagoreans, the explanation of the musical faculty lay in the mathematics of integer ratios and the physics of simply vibrating strings. Helmholtz was able to draw upon nineteenth century physics, for a more properly mechanistic and complete explanation of the phenomenon of consonance. For him, a mechanism was a physical device such as a real resonator or oscillator. The principal tool that we have available, beyond those that Aristotle and Helmholtz knew of, is the computer.

Of course, it is often the algorithm that the computer executes that is of interest, rather than the computer itself, since for many interesting cases we can state the algorithm independently of any particular machine. However, the idea of an algorithm is not in itself novel. Algorithms (such as Euclid's algorithm) were known to Helmholtz. It is the computer which transforms the notion of an algorithm from a procedure that needs a person to execute it to the status of a mechanism or explanation.

## 2. Consonance

Helmholtz (1862) explained the dimension of Consonance in terms of the coincidence and proximity of the overtones and difference tones that arise when simultaneously sounded notes excite real non-linear physical resonators, including the human ear. To the extent that an interval's most powerful secondary tones exactly coincide, it is consonant or sweet-sounding. To the extent that any of its secondaries are separated in frequency by a small enough difference to "beat" at a rate which Helmoltz puts at around 33 *c/s*, it is dissonant, or harsh. Thus for the diatonic semitone, with a frequency ratio of 16/15, only very high, low-energy overtones coincide, so it is weakly consonant, while the two fundamentals themselves produce beats, in the usual musical ranges, so it is also strongly dissonant. For the perfect fifth, on the other hand, with a frequency ratio of 3/2, all its most powerful secondaries coincide, and only very weak ones are close enough to beat. The fifth is therefore strongly consonant and only weakly dissonant. This theory, which has survived (with an important modification due to Plomp and Levelt 1965) to the present day, successfully explains not only the subjective experience of consonance and

dissonance in chords, and the effects of chord inversion, but also the possibility of Equal Temperament. The latter is the trick whereby by slightly mistuning all the semitones of the octave to the same ratio of $\sqrt[12]{2}$, one can make an instrument sound tolerably in tune in all twelve major and minor keys. Equal Temperament distorts the seconds and thirds (and their inverses the sevenths and sixths) more than the fourths and the fifths, and affects the octaves hardly at all. Helmholtz' theory predicts than distortion to the seconds and thirds will be less noticeable that distortion to the latter, so it explains why this works.

However, Helmholtz recognised very clearly that this success in explaining equal temperament raised a further question which his theory of consonance could not answer, namely what it is that makes the character of an augmented triad (C E G♯) or a diminished seventh chord (C E♭ G♭ B♭♭ ) so different from that of a major or minor triad. Consonance does not explain this effect, since all four chords when played on an equally-tempered instrument are entirely made up of minor and major thirds. He correctly observes that one of the equally-tempered major thirds in the augmented triad is always heard as the harmonically remote diminished fourth, and observes that "this chord is well adapted for showing that the original meaning of the intervals asserts itself even with the imperfect tuning of the piano, and determines the judgement of the ear." (Cf. Helmholtz 1862, as translated by Ellis 1885, p.213 and cf. p.338). But Helmholtz had no real explanation for how this could come about.

It is in no way to Helmholtz' discredit that this was so. He did in fact sketch an answer to the problem, and it is striking that his way of tackling it is essentially algorithmic, despite the fact that it implies a class of mechanism that he simply did not have a way of reifying. However, Helmholtz tried to approach the perceptual effect as one of dissonance, while in reality it concerns an entirely orthogonal relation between notes, namely the one that musicians usually refer to as the "harmonic" relation. This relation, which underlies phenomena like chord progression, key, and modulation, is quite independent of consonance, although both have their origin in the Pythagorean integer ratios.

## 3. Harmony

The first complete formal identification of the nature of the harmonic relation is in Longuet-Higgins (1962a, 1962b) (cf. the paper in this volume), although there are some earlier incomplete proposals, including work by Weber, Schoenberg, Hindemith, and the important work of Ellis (1874, 1875), to which we return below. Longuet-Higgins showed that the set of musical intervals relative to some fundamental frequency was the set of ratios definable as the product of powers of the prime factors 2, 3, and 5, and no others – that is as a ratio of the form $2^x.3^y.5^z$, where $x$, $y$, and $z$ are positive or negative integers. (The fact that ratios involving factors of seven and higher primes do not contribute to this definition of harmony does not exclude them from the theory of consonance. In real resonators, overtones involving such factors do arise, and contribute to consonance. Helmholtz realised that the absence of such ratios from the chord system of tonal harmony represented a problem for his theory of chord function, and attempted an explanation in terms of consonance – cf. Ellis (translation) 1885, p.213-213).

Longuet-Higgins' observation means that the intervals form a three-dimensional discrete space, with those factors as its generators, in which the musical intervals can be viewed as vectors. Since the ratio 2 corresponds to the musical octave, and since for most harmonic purposes, notes an octave apart are functionally equivalent, it is convenient to project the three dimensional space along this axis into the 3 x 5 plane. It then appears as in

Figure 1, adapted from Longuet-Higgins (1962a), in which the (not terribly systematic) traditional interval names are associated with positions in the plane. As Longuet-Higgins

| | | | | | | |
|---|---|---|---|---|---|---|
| Augmented Seventh | Augmented Fourth | Small Half-Tone | Augmented Fifth | Augmented Second | Augmented Sixth | Augmented Third |
| Imperfect Fifth | Minor Tone | Major Sixth | Major Third | Major Seventh | Tritone | Small Limma |
| Imperfect Third | Dominant Seventh | Perfect Fourth | Unison | Perfect Fifth | Major Tone | Imperfect Sixth |
| False Octave | Minor Fifth | Semitone | Minor Sixth | Minor Third | Minor Seventh | Imperfect Fourth |
| Diminished Sixth | Diminished Third | Diminished Seventh | Diminished Fourth | Diminished Octave | Diminished Fifth | Great Limma |

Figure 1: The space of harmonic intervals (adapted from Longuet-Higgins 1962a).

points out, the musician's notion of harmonic distance or "remoteness" of intervals is very directly reflected by a number of simple metrics upon this space, of which the summed "city block" distance between points is the most obvious, while the minimum spanning rectangle is another.

It is convenient to represent the space in terms of the traditional note-names that would be associated with each of these intervals relative to an origin of C, as in Figure 2. The note names are ambiguous with respect to the intervals, and the entire space now repeats itself in a south-easterly direction. (That is to say that the note names "wrap" the full space onto a cylinder, which is here projected back onto the plane). While we generated this map from an origin of C, any of the positions can now be regarded as the origin: if we slide the earlier interval-name space (Figure 1) across the note-name space (Figure 2), the former will correctly identify the note name for all intervals from any origin.

The distortions of equal temperament have the effect not only of equating pairs of frequencies with the same name, such as the various Cs in the figure, but also pairs such as G♯ and A♭. That is to say that equal temperament maps the note-name space into a torus, with twelve positions on it. If we project this highly ambiguous set onto the full plane, using the numbers 0 (for C, B♯, etc) to 11 (for B, C♭, etc) for the twelve notes of the equally-tempered octave, it looks like Figure 3.

Helmholtz' problem can now be formulated as follows in terms of Longuet-Higgins' theory. When we hear an equally-tempered chord, we project each ambiguous equally-tempered note onto all possible interpretations in some portion of the full space, relative to some origin. (A sensible interpretation of "some portion" would be the region defined by the traditional interval names, Figure 1). We then pick one interpretation for each note, on the basis of one of our two metrics. With the major and minor triads, there is a way of picking a single interpretation for each note that makes all intervals between pairs of notes in the chord a major or minor third, or a perfect fifth (or their inverses). The problem can be visualised as in Figure 4, in which it will be apparent that there are several such clusters, all equivalent under translation. The same applies to the minor triad, as

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $G\sharp$ | $D\sharp$ | $A\sharp$ | $E\sharp$ | $B\sharp$ | $Fx$ | $Cx$ | $Gx$ | $Dx$ |
| $E$ | $B$ | $F\sharp$ | $C\sharp$ | $G\sharp$ | $D\sharp$ | $A\sharp$ | $E\sharp$ | $B\sharp$ |
| $C$ | $G$ | $D$ | $A$ | $E$ | $B$ | $F\sharp$ | $C\sharp$ | $G\sharp$ |
| $A\flat$ | $E\flat$ | $B\flat$ | $F$ | $C$ | $G$ | $D$ | $A$ | $E$ |
| $F\flat$ | $C\flat$ | $G\flat$ | $D\flat$ | $A\flat$ | $E\flat$ | $B\flat$ | $F$ | $C$ |
| $D\flat\flat$ | $A\flat\flat$ | $E\flat\flat$ | $B\flat\flat$ | $F\flat$ | $C\flat$ | $G\flat$ | $D\flat$ | $A\flat$ |
| $B\flat\flat\flat$ | $F\flat\flat$ | $C\flat\flat$ | $G\flat\flat$ | $D\flat\flat$ | $A\flat\flat$ | $E\flat\flat$ | $B\flat\flat$ | $F\flat$ |

Figure 2: The space of note-names (adapted from Longuet-Higgins, 1962a).

the reader may easily verify. However, the augmented chord that caused Helmholtz such trouble does not have this property. All ways of selecting a single interpretation for all three notes force one of the equally-tempered major thirds to be interpreted as a more remote augmented/diminished interval, as can be seen in Figure 5.

The augmented chord differs from the major and minor triads in another way. Whereas examination of Figure 4 will show that, once a particular C has been chosen, there is a unique closest cluster of interpretations for the other notes, this is not true for the augmented chord in Figure 5. C E G♯ is no more and no less closely grouped than C E A♭. The interpretation remains ambiguous until we hear the following chord, which "resolves" the ambiguity. For example, if this chord is an F major triad, then we hear the ambiguous chord as the first alternative. This resolution is strongly influenced by progressions of a semitone between notes in the first chord and the second, as is shown by the fact that the resolution in question is considerable reinforced if the dominant seventh note B♭ is added to the augmented chord (by constrast, a resolution onto a D♭ major triad is not particularly convincing). These claims can be verified by inspecting Figure 6. It is important to note that the cluster of interpretations that results for the "augmented plus seventh" chord is not a unique tightest cluster under either of the metrics that were mentioned earlier. Under the minimal spanning rectangle metric it is among an equivalence class of minimal clusters. Under the alternative city block metric, it is not even minimal, although it is not far off. This is an indication that in interpreting a chord, we will in general need to take its context, and particularly the succeeding chord into

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 8 | 3 | 10 | 5 | 0 | 7 | 2 | 9 | 4 |
| 4 | 11 | 6 | 1 | 8 | 3 | 10 | 5 | 0 |
| 0 | 7 | 2 | 9 | 4 | 11 | 6 | 1 | 8 |
| 8 | 3 | 10 | 5 | 0 | 7 | 2 | 9 | 4 |
| 4 | 11 | 6 | 1 | 8 | 3 | 10 | 5 | 0 |
| 0 | 7 | 2 | 9 | 4 | 11 | 6 | 1 | 8 |
| 8 | 3 | 10 | 5 | 0 | 7 | 2 | 9 | 4 |

Figure 3: The space of Equal Temperament (adapted from Longuet-Higgins & Steedman, 1971).

account.

All of these characteristics hold of the diminished chord C E♭ G♭ B♭♭: again, verifying this fact is suggested as an exercise.

It is interesting to ask at this point why the tonal harmonic space should involve the three dimensions associated with prime factors of two, three, and five. One might imagine that the answer might be physiological, or even that this fact might be an accident. However, it turns out that the answer is again essentially algorithmic. It is easy to see that this particular space is the unique highest dimensional space in which positions which are close in frequency (and therefore confusable to an ear with limited acuity) are widely separated, and therefore can be disambiguated by context in the manner just discussed. For example, the inclusion of the ratio seven, which intruduces a note close in frequency to the dominant seventh (see Figure 1), at a remove of only three steps in harmonic space from the true dominant seventh. By contrast the spatial distance between the major and minor tones is five steps. It follows that while musics based on other ratios can be constructed (and probably have arisen naturally), and can be perfectly consonant, they are necessarily more restricted harmonically. In particular, they can have no equivalent of equal temperament, and no scope for the richness of harmonic development that it permits. (That is not of course to imply that such musics are less interesting than tonal music, merely that they must achieve their richness on some other dimension – for

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| . | . | . | . | 0 | 7 | . | . | 4 |
| 4 | . | . | . | . | . | . | . | 0 |
| 0 | 7 | . | . | $E$ | . | . | . | . |
| . | . | . | . | $C$ | $G$ | . | . | 4 |
| 4 | . | . | . | . | . | . | . | 0 |
| 0 | 7 | . | . | 4 | . | . | . | . |

Figure 4: The projection of an equally-tempered chord of C major.

example, in rhythm).

## 4. Algorithms and Computational Architectures

Although the above discussion has referred to processes of searching and clustering, we have not yet said anything about how these computations might be carried out. The algorithm implicit in the above examples is to map the torus of equal temperament onto a suitably circumscribed portion of the plane, and serially compute for chords and chord sequences the tightest cluster(s) containing one interpretation for each equally-tempered note. This is the tactic discussed in Steedman (1973). However, this process is also parallellisable, and one way to think about it is to think in terms of a neural net, with inputs corresponding to the twelve degrees of the equally-tempered scale, and a considerably larger number of outputs corresponding to interpreted diads, triads, and so on, each associated with a set of fully disambiguated positions in the full space of just intonation. This device would be a close relative of the approach of Bharucha (1987, cf. Jones et al. 1988), differing only in having a considerably greater variety of chord units, and in mapping those units onto Longuet-Higgins' harmonic representation, and is being investigated by Dan Petit at the University of Pennsylvania. (Such nets could conceivably be trained by one of the standard algorithms. However, it seems more likely that human novices add new chord units piecemeal, covering a larger and larger region of the harmonic space.)

The use of such a device is not quite as straightforward as the above remarks imply. The harmonic centre of a piece may move via the chord of the mediant or major third and an extended sequences of descending fifths (that is, via an extended "perfect cadence"), to a *different* instance of its supposed tonic. (Examples are afforded by the opening *tutti*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 8 | . | . | . | 0 | . | . | . | 4 |
| 4 | . | . | . | 8 | . | . | . | 0 |
| 0 | . | . | . | 4 | . | . | . | 8 |
| 8 | . | . | . | 0 | . | . | . | 4 |
| 4 | . | . | . | 8 | . | . | . | 0 |
| 0 | . | . | . | 4 | . | . | . | 8 |

Figure 5: The projection of an equally-tempered augmented chord

of Beethoven's fourth piano concerto in G, and by *Basin Street Blues*. The latter does this trick repeatedly, exploiting the perfect-cadence -inducing dominant seventh chord extensively, in a manner discussed at length in Steedman 1984. This fact lends this piece a feeling of perpetual paradoxical motion, exemplified in a famous recording by Louis Armstrong.) For this reason, any fixed finite net must be mapped onto the cylindrical space of traditional notation. In terms of the model involving piecemeal addition of units, one must envisage a developmental stage at which the novice recognises that his or her harmonic space can be wrapped into a cylinder. Such a mapping arguably preserves all the information in the interpretation that a musician would regard as significant.

A similar perpetual motion in the vertical direction, along the major third axis, does not seem to be nearly as compelling. The reason is presumably that there can be no really convincing "mediant-cadential" chord equivalent to the dominant seventh chord. (This in fact follows from the characteristics of Longuet-Higgins' space, and the fact that confusable intervals like the augmented fifth and diminished fourth are not widely separated on this axis.) It is therefore not necessary to map such networks onto the torus, as in Bharucha's model of key identification. Nor is this move desirable, since for the purpose of identifying the harmonic function of chords and notes within chords, the tactic loses information that a musician would regard as significant.

Which style of algorithm we use does not greatly matter, and in fact the net representation can be regarded as a compiled form of the serial algorithm, derivable by network learning techniques. What is more important is to recall that chord-based clustering alone is not enough to disambiguate chord function, as we saw in the case of the augmented and diminished chords, and as is in fact the case with virtually all chords except the minor and major triads, and the major seventh chord, all of which are extremely resolved. (It is probably even possible to contextualise the minor triad so that it is perceived as including

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 8 | . | 10 | . | 0 | . | . | . | 4 |
| 4 | . | . | . | $\swarrow G\sharp$ | . | 10 | . | 0 |
| 0 | . | . | $(A)$ | $\swarrow E$ | . | . | . | 8 |
| 8 | . | $B\flat\nearrow$ | $(F)$ | $C$ | . | . | . | 4 |
| 4 | . | . | . | 8 | . | 10 | . | 0 |
| 0 | . | . | . | 4 | . | . | . | 8 |

Figure 6: The projection of an equally-tempered augmented seventh chord.

an augmented second rather than a minor third, although most styles of tonal music will collapse under the strain). We usually need to look at the succeeding chord to decide which interpretation is correct.

Although Bharucha extends his net-based analyser to deal with sequences of chords and tonalities, and applies it to the task of identifying the key of such sequences, it is not entirely clear that net-based parallel techniques, which integrate over wide stretches of music with rather indefinite boundaries, are really appropriate for the task of interpreting chord sequences. This is particularly likely to be the case in assigning tonalities or chordal accompaniments to melodies, in which the transitions between tonalities which contribute to the identification of key seem to be quite abrupt and all-or-none in character.

## 5. Key-analysis in Unaccompanied Melodies

Consider a listener who hears an unaccompanied melody for the first time. A minimal requirement for us to agree that they have correctly understood the piece is for them to be able to identify the kinds of harmonic relationships that are implicit in the key signature. We can translate this into the problem of correctly identifying the position in Longuet-Higgins space for the interpretation of each note. Of course, most listeners will not have perfect pitch, so we will allow them to do this relative to an arbitrary origin, such as the first note of the piece. But we shall insist that they correctly identify the key note, or at least the sequence of tonalities involved. (Of course, most listeners do not have the vocabular to identify these properties either, but we can show by getting them to perform various completion tasks and error detection tasks that everybody has this knowledge implicitly).

To discuss this problem concretely we need a corpus of melodies. Since we have been

discussing equal temperament, we will follow Longuet-Higgins and Steedman (1971) in choosing the subjects of the fugues from Bach's Well-tempered Keyboard. Not least among the virtues of this corpus is the fact that it has not been assembled with any particular theory of processing in mind. However, it has one peculiarity that it would be wrong to take advantage of: all the subjects happen to begin on either the tonic or the dominant. Since this is not a characteristic of tonal melodies in general – nor even of Bach's fugue subjects – we shall eschew any rules that exploit this fact. (Thus we avoid the infamous "tonic-dominance preference rule" of the earlier paper entirely).

The first point to note is that merely identifying the tightest clustering of interpretations of the earliest notes in the piece will not yield a correct identification of its key. While it may be reasonable to believe that a piece will not include an imperfect interval until the key has been sufficiently resolved for it to be clear that it *is* imperfect, we can in the case of the A minor fugue of book I (Figure 7) find a subject which has a diminished seventh as its third interval. The tightest cluster of interpretations for the first four notes under

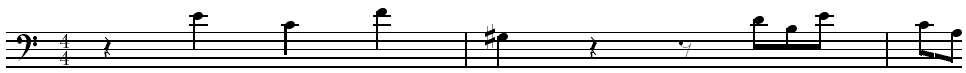

Figure 7: A minor fugue, book II.

the city block metric (though not under the minimal spanning rectangle) is one in which the G♯ is interpreted as an A♭, implying a key of F minor. However, no human listener would make this mistake.

At the other extreme, when the tonality is clearly and unambiguously established, say by a major or minor arpeggio, then it is virtually impossible to entertain the hypothesis that any of the intervals is imperfect. For example, it is hard to hear the E minor subject of book I as being in the key of G♯ minor, and to interpret the G as an F double-sharp, despite the fact that the first six notes are all compatible with the latter key, while the seventh note is an accidental in both. We cannot escape hearing the first four notes as a chord of E minor (Figure 8).

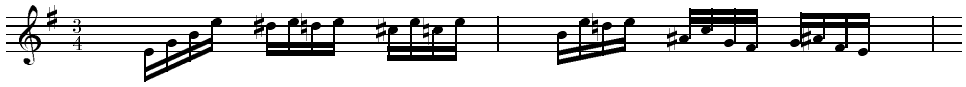

Figure 8: E minor fugue, book I.

However, to identify the initial tonality is not the same as identifying the key. Even if the piece begins with the notes of a major or minor triad, this initial tonality may be part of a cadence *onto* the tonic, rather than the tonic itself. The D major subject from book II, Figure 8, is a case in point. The initial tonality here is undoubtedly that of G major, the subdominant of the key as Bach wrote it, and indeed up to the fifth note the key could perfectly well be Gmajor. ('La Cucaracha' is a conveniently well-known example of a melody which begins in essentially the same way, give or take an octave, and for which the hypothesis that the repeated initial note was the tonic would be quite correct). It is only after the transition from the tied-note on B to an E (another instance of an early imperfect interval) and the succeeding A, establishing a new tonality of the dominant A

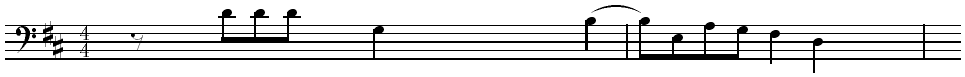

Figure 9: D major fugue, book II.

major, that we suspect what is confirmed by the subsequent F♯ and D, namely that this is a IV,V,I cadence onto the tonic D.

But how do we know that? Since there is no C or C♯ anywhere in the melody, we could in fact notate this example in G, with the imperfect fourth falling between the E and the A, rather than the B and the E. As in the case of the C♯ minor fugue, Figure 14, we seem to require an analysis of the piece at a higher level than mere individual notes. In fact, we need something that it is tempting to call a *grammar* of melody, whose syntax captures such structures as the repeated initial note and the scale progression from the A to the F♯ as structural *constituents*, and whose semantics defines interpretations of such constituents in terms of the harmonic space. As we have already noted, this kind of fine-grained analysis is something that neural nets seem quite ill-adapted to. The lack of such an analysis in terms of chord progressions rather than global properties of the melody was also the major shortcoming of the otherwise closely related approach in Steedman 1973.

## 6. Towards a Grammar of Melodic Tonality

The work of Lerdahl and Jackendoff (1983) at the structural end of such grammars, drawing on the Chomskean tradition of Generative Grammar, and of Narmour (1977, 1990) building on the more psychologistic approach of Meyer in a more interpretative direction, is particularly important. I think it is fair to say, however, that such frameworks provide both more and less than we need to solve the problem that Helmholtz bequeathed to us. They provide more in the sense that the structures that they encompass are far more extensive than those that we need for the local analysis of tonality. They provide less than we need in the sense that the link between structural rules and interpretative rules – that is, the equivalent of a semantics – is as yet somewhat underspecified.

Whatever the limitations on our access to natural language semantics, (cf. Chomsky 1957), the study of its syntax would not have got far if it had not been informed by some fairly strong intuitions about meaning. While recent work by these authors and their colleagues is explicitly addressed to this question, and appears extremely promising (I am thinking particularly of Lerdahl 1988, (who discusses a number of related approaches to harmony, including those of Balzano 1982, Shepherd 1982, and Krumhansl and Kessler 1982), and of Narmour 1992, and references therein). I think it may in the meantime be worth sketching the form that such a grammar might take if we were to assume (in the Fregean tradition of Montague 1974) that the structural rules of such a grammar should be related as closely and simply as possible to rules of interpretation. The earlier example of the D major fugue (Figure 9) and many others among the subjects of the Forty-eight show that such a semantics must compositionally define key in terms of cadences, or progressions of chord-tonalities, perhaps along the lines suggested in Steedman (1984), with positions in the harmonic space playing much the same role as individuals in the "model" in linguistic semantics, and with the rules defining cadences playing much the same role as rules of logical inference.

The earlier remarks suggest that our grammar of melody should be concerned with movements between points in the harmonic space, and their relation to basic major and minor triads, or tonalities, as in Steedman (1973). We have already seen in the case of the D major fugue, Figure 9, that more than one successive note may correspond to the same position, and that a movement may either be a discontinuous jump, or a scale movement. Thus the three repeated eighth-note Ds at the start of this subject appear to be more or less equivalent to a single dotted quarter-note D, and the scale progression from A to F♯ in the middle of the second bar seems more or less equivalent to a jump from the former to the latter. More interestingly, sequences of different notes may count as staying in the same place in harmonic terms, and sequences of notes separated by intervals other than seconds may count as scale transitions between harmonic positions.

In the first category, various kinds of trills and twiddles can be equivalent to a single note. For example, the E♭ major subject of Book I is heard as beginning with a tonic arpeggio, in which the mediant is realised as an "inflection" consisting of three notes, of which only the first and last are actually G (Figure 10). A similar configuration

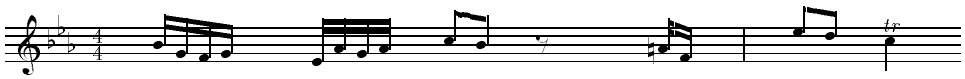


Figure 10: E♭ major fugue, book I.

immediately following is also perceived as equivalent to A♭. There are a number of such configurations of notes of the same duration separated by upward and downward seconds that have the same effect.

It is convenient to refer to such sequences, which count as equivalent to a single harmonic position, as "points". Such configurations have a recursive character. For example, we have seen that repeated notes of the same pitch and duration count as a point of the same total duration. It is also the case that a "turn" of notes of the same duration and the following configuration of ascending and descending seconds can be heard as a point whose value or effective pitch is that of the repeated note, as in the C♯ major subject of Book II, Figure 11. The G major fugue, book I, Figure 12, suggests that a note followed

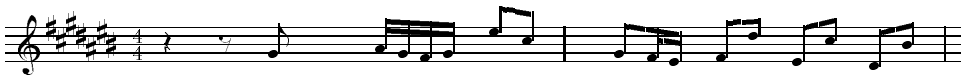

Figure 11: C♯ major fugue, book II.

by a turn on the same note also counts as a point. This kind of recursive constituency is

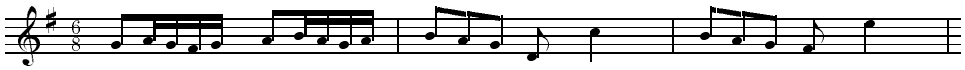

Figure 12: G major fugue, book I.

familiar from phrase-structure grammars for natural languages, and it seems possible in principle to construct more deeply embedded examples.

Similarly, just as there is more than one way of staying in the same place, so there is more than one way of getting from one place to another. We have already seen that a scale of seconds of the same duration and direction is equivalent to a discontinuous jump between its endpoints. Scale movements also can be recursive in character, either by involving complex points rather than individual notes, or by virtue of interleaving scales in parallel motion, or by interleaving a repeated same pitch, as in the case of the E minor fugue, Book I, Figure 8.

The construction of a parser for such a grammar is quite a difficult task. The definitions of points, including inflections and turns, and the definition of scales, in terms of ascending and descending seconds of the same duration are locally ambiguous. For example, how are the sequences in Figures 13, 14, and 15 to be "parsed" according to these grammars? Clearly, the answer to this question depends on our perception of the metric structure

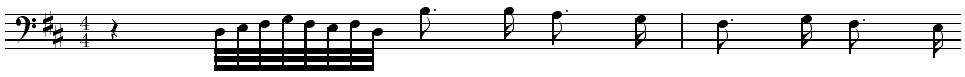

Figure 13: D major fugue, book I

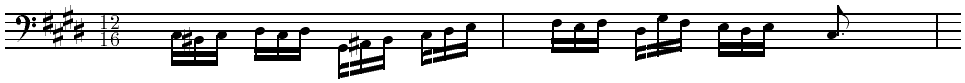

Figure 14: C ♯ minor fugue, book II

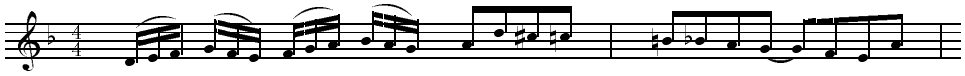

Figure 15: D minor fugue, book II

of these pieces, as indicated in the time signature, bar lines etc. Thus the three pieces seem to be heard as in Figures 16, 17, and 18.

The perception of rhythm and metre has been investigated by Longuet-Higgins (1994), and references therein, and by Steedman (1977). However, the integration of metrical and harmonic analysis of the kind that is required for a really adequate account of key analysis has hardly begun, and seems quite challenging, as does the integration of notions of cadence and chord progression. For example, the question of whether the key of the C major fugue in Book I (Figure 19) is in fact C, rather than F, and whether it therefore begins with a tonality of IV or of I, rests on the question of whether it is the interval between the A and D of notes 8 and 9, or that between the same D and the succeeding G that is imperfect. This decision seems to rest upon the considerable rhythmical salience of the syncopated G itself: an imperfect interval onto such a resting place seems unlikely. Nor will clustering do anything for us here: the clusters under the two key analyses are identically close, under any conceivable metric.

The powerful theories of Lerdahl and Jackendoff, and Narmour, and the references

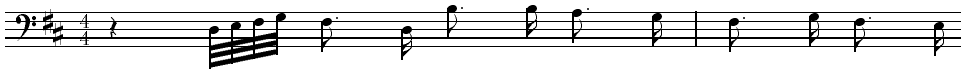

Figure 16: Cf. Figure 14.

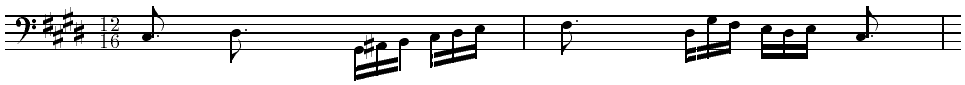

Figure 17: Cf. Figure 15.

already cited, will undoubtedly be as helpful in this venture as will as the constructive methods of the computer scientists that I have concentrated on here.

## 7. Conclusion

What has been presented in this paper is work in progress by a number of scientists in a number of disciplines. The problem that they are trying to solve is a difficult one, and the solutions remain incomplete. In the terms of the question that is addressed at this conference, as suggested in its title, it would be premature to claim a "new breakthrough". On the other hand, they do not seem to be a dead end. The computer has already provided an entirely new kind of algorithmic answer to questions about the nature of mind, which it is simply impossible to imagine having to do without.

In pursuit of this argument, I would like to return for a moment to the question of why Helmholtz did not manage to answer his own beautifully simple question concerning the nature of our experience of equal temperament.

Helmholtz actually had access to more of the crucial concepts that were needed for an answer than I have so far revealed. A very close relative of Longuet-Higgins' harmony theory was available during Helmhotz's lifetime. In fact it was presented to this Society, in a paper by Ellis (1874), entitled 'On Musical Duodenes', concerning the nature of modulation. We know that Helmholtz at least had access to this work, for the following curious reason. The translator of Helmholtz' 1862 book was none other than Ellis (1875), who greatly expanded the original by the addition of numerous appendices, mostly concerning a variety of novel keyboard instruments and tables of the precise frequencies of the pipes in the organs of the more significant churches of Europe – a fact of which we know that Helmholtz was aware, since he took exception to these rather extensive additions.

One of these appendices consisted of a fairly complete version of his paper on modulation to the Royal Society of the previous year, including the diagram reproduced in Figure 20 (taken from the second edition of Ellis' translation 1885, p.463, where he gives references to related even earlier work by Weber.).

We shall of course probably never know whether Helmholtz got as far as actually reading Appendix XX of Ellis' translation. But it is striking that neither he, nor Ellis, not any of their contemporaries, seem to have seen that this diagram, which is in essence a reflection and a rotation of that proposed by Longuet-Higgins, needed only the notion of computation to breathe it into life as an answer to the question that Helmholtz had so clearly recognised.

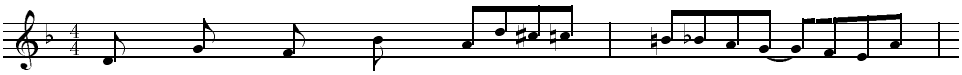

Figure 18: Cf. Figure 16.

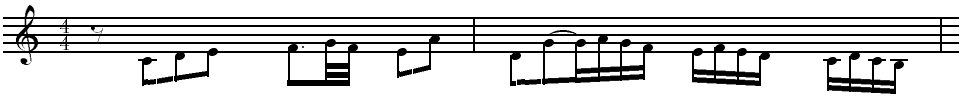

Figure 19: C major fugue, book I.

| | | | | | | |
|---|---|---|---|---|---|---|
| $F\flat$ | $A\flat$ | $C$ | $E$ | $G\sharp$ | $B\sharp$ | $Fx$ |
| $B\flat\flat$ | $D\flat$ | $F$ | $A$ | $C\sharp$ | $E\sharp$ | $Gx$ |
| $E\flat\flat$ | $G\flat$ | $B\flat$ | $D$ | $F\sharp$ | $A\sharp$ | $Cx$ |
| $A\flat\flat$ | $C\flat$ | $E\flat$ | $G$ | $B$ | $D\sharp$ | $Fx$ |
| $D\flat\flat$ | $F\flat$ | $A\flat$ | $C$ | $E$ | $G\sharp$ | $B\sharp$ |
| $G\flat\flat$ | $B\flat\flat$ | $D\flat$ | $F$ | $A$ | $C\sharp$ | $E\sharp$ |
| $C\flat\flat$ | $E\flat\flat$ | $G\flat$ | $B\flat$ | $D$ | $F\sharp$ | $A\sharp$ |
| $F\flat\flat$ | $A\flat\flat$ | $C\flat$ | $E\flat$ | $G$ | $B$ | $D\sharp$ |
| $B\flat\flat\flat$ | $D\flat\flat$ | $F\flat$ | $A\flat$ | $C$ | $E$ | $G\sharp$ |

Figure 20: The Duodenarium (adapted from Ellis 1874, 1885)

**Acknowledgements**

I am grateful to Eric Clarke for extensive comments on the draft, and to Peter Buneman, Dan Hardt, Dan Petit, and Scott Prevost for recent discussions on these questions, and to Christopher Longuet-Higgins for advice over a longer period. The research was supported in part by NSF grant nos. IRI90-18513, IRI90-16592, IRI91-17110 and CISE IIP-CDA-88-22719, DARPA grant no. N00014-90-J-1863, and ARO grant no. DAAL03-89-C0031.

**Bibliography**

Balzano, G. 1982 The Pitch Set as a Level of Description for Studying Musical Perception, in M. Clynes (ed.), Music, Mind and Brain, New York, Plenum.

Bharucha, J. 1987 Music Cognition and Perceptual Facilitation: a Connectionist Framework. *Music Perception* **5**, 1-30

Chomsky, N. 1957 *Syntactic Structures*, Mouton, The Hague.

Ellis, A. 1874 On Musical Duodenes. *Proceedings of the Royal Society,* **23**,

Ellis, A. 1875 Appendix XX to the 1st ed. of his translation of Helmholtz 1862 (q.v.), as *On the Sensations of Tone*, Longmans.

Ellis, A. 1875 2nd ed. of his translation of Helmholtz 1862, as *On the Sensations of Tone*, Longmans, republished by Dover, New York, 1954.

Helmholtz, H. 1862 *Die Lehre von dem Tonempfindungen*, Vieweg, Braunschweig.

Jones, J. Miller, B. & Scarborough, D. 1988 A Rule-based Expert System for Music Perception. *Behaviour Research Methods, Instruments and Computers,* **20**, 255-262.

Krumhansl, Carol, and Kessler, E. 1982 Tracing the dynamic changes in perceived tonal organisation in a spatial representation of musical keys. *Psychological Review,* **89**, 334-368.

Lerdahl, F. & Jackendoff, R. 1983 *A Generative Theory of Tonal Music*, MIT Press, Cambridge MA.

Lerdahl, F. 1988 Tonal Pitch Space. *Music Perception,* **5**, 315-350.

Longuet-Higgins, H.C. 1962a Letter to a Musical Friend. *The Music Review,* **23**, 244-248.

Longuet-Higgins, H.C. 1962b Second Letter to a Musical Friend. *The Music Review,* **23**, 2271-280.

Longuet-Higgins, H.C. 1994 This conference: incomplete reference.

Longuet-Higgins, H.C. & Steedman, M. 1971 On Interpreting Bach, in B. Meltzer and D. Michie, (eds.) *Machine Intelligence* **6**, Edinburgh University Press.

Montague, R. 1974 *Formal Philosophy*, ed. R. Thomason, New Haven, Yale University Press

Narmour, E. 1977 *Beyond Shenkerism*, Chicago, Chicago University Press.

Narmour, E. 1990 *The Analysis and Cognition of Basic Musical Structures*, Chicago, Chicago University Press.

Narmour, E. 1992 *The Analysis and Cognition of Musical Complexity*, Chicago, Chicago University Press.

Plomp, R. & Levelt, W. 1965 Tonal Consonance and Critical Bandwidth. *Journal of the Acoustical Society of America,* **38**, 548-560

Shepard, R. 1982 Structural representations of musical pitch, in Deutsch, D. (ed.), *The Psychology of Music*, New York, Academic Press.

Steedman, M. 1973 *The Formal Description of Musical Perception*, unpublished Ph.D dissertation, University of Edinburgh.

Steedman, M. 1977 The Perception of Musical Rhythm and Metre. *Perception,* **6**, 555-569.

Steedman, M. 1984 A Generative Grammar for Jazz Chord Sequences. *Music Perception,* **2**, 52-77.