Set-Class Similarity, Voice Leading, and the Fourier Transform

Dmitri Tymoczko

Abstract   In this article, I consider two ways to model distance (or inverse similarity) between chord types, one based on voice leading and the other on shared interval content. My goal is to provide a contrapuntal reinterpretation of Ian Quinn's work, which uses the Fourier transform to quantify similarity of interval content. The first section of the article shows how to find the minimal voice leading between chord types or set-classes. The second uses voice leading to approximate the results of Quinn's Fourier-based method. The third section explains how this is possible, while the fourth argues that voice leading is somewhat more flexible than the Fourier transform. I conclude with a few thoughts about realism and relativism in music theory.

TWENTIETH-CENTURY MUSIC OFTEN MOVES FLEXIBLY between contrasting harmonic regions: in the music of Stravinsky, Messiaen, Shostakovich, Ligeti, Crumb, and John Adams, we find diatonic passages alternating with moments of intense chromaticism, sometimes mediated by nondiatonic scales such as the whole-tone and octatonic. In some cases, the music moves continuously from one world to another, making it hard to identify precise boundaries between them. Yet we may still have the sense that a particular passage, melody, or scale is, for instance, fairly diatonic, more-or-less octatonic, or less diatonic than whole-tone. A challenge for music theory is to formalize these intuitions by proposing quantitative methods for locating musical objects along the spectrum of contemporary harmonic possibilities.

One approach to this problem uses voice leading: from this point of view, to say that two set-classes are similar is to say that any set of the first type can be transformed into one of the second without moving its notes very far. Thus, the acoustic scale is similar to the diatonic because we can transform one into the other by a single-semitone shift; for example, the acoustic scale {C, D, E, F♯, G, A, B♭} can be made diatonic by the single-semitone displacement F♯ → F or B♭ → B. Similarly, when we judge the minor seventh chord

to be very similar to the dominant seventh, we are saying that we can relate them by a single-semitone shift. This conception of similarity dates back to John Roeder's work in the mid-1980s (1984, 1987) and has been developed more recently by Thomas Robinson (2006), Joe Straus (2007), and Clifton Callender, Ian Quinn, and myself (2008). The approach is consistent with the thought that composers, sitting at a piano keyboard, would judge chords to be similar when they can be linked by small physical motions.

Another approach uses intervallic content: from this point of view, to say that set-classes are similar is to say that they contain similar collections of intervals. (That the two methods are different is shown by "Z-related" or "nontrivially homometric" sets, which contain the same intervals but are nonidentical according to voice leading.) In a fascinating pair of papers, Quinn has demonstrated that the Fourier transform can be used to quantify this approach.[1] Essentially, for any number n from 1 to 6, and every pitch class p in a chord, the Fourier transform assigns a two-dimensional vector whose components are

$$V_{p,n} = (\cos 2\pi pn/12, \sin 2\pi pn/12). \tag{1}$$

Adding these vectors together, for one particular n and all the pitch classes p in the chord, produces a composite vector representing the chord as a whole—its "nth Fourier component." The length (or "magnitude") of this vector, Quinn astutely observes, reveals something about the chord's harmonic character: in particular, chords saturated with $(12/n)$-semitone intervals, or intervals approximately equal to $12/n$, tend to score highly on this index of chord quality.[2] The Fourier transform thus seems to capture the intuitive sense that chords can be more or less diminished-seventh-like, perfect-fifthy, or whole-tonish. It also seems to offer a distinctive approach to set-class similarity: from this point of view, two set-classes can be considered "similar" when their Fourier magnitudes are approximately equal—a situation that obtains when the chords have approximately the same intervals.

The interesting question is how these two conceptions relate. In recent years, a number of theorists have tried to reinterpret Quinn's Fourier magnitudes using voice-leading distances. Robinson (2006), for example, pointed out that there is a strong anticorrelation between the magnitude of a chord's first Fourier component and the size of the minimal voice leading to the nearest chromatic cluster. (See also Straus 2007, which echoes Robinson's point.) However, neither Robinson nor Straus found an analogous interpretation of the other Fourier components. In an interesting article in this issue (see pages 219–49), Justin Hoffman extends this work, interpreting Fourier components in light of unusual "voice-leading lattices" in which voices move by distances other than one semitone. But despite this intriguing idea, the

---

1 See Quinn 2006 and 2007. Quinn's use of the Fourier transform develops ideas in Lewin 1959 and 2001 and Vuza 1993.

2 These magnitudes are the same for transpositionally or inversionally related chords, so it is reasonable to speak of a *set-class's* Fourier magnitudes.

relation between Fourier analysis and more traditional conceptions of voice leading remains obscure.

The purpose of this article is to describe a general connection between the two approaches: it turns out that the magnitude of a chord's *n*th Fourier component is approximately inversely related to the size of the minimal voice leading to *the nearest subset* of any perfectly even *n*-note chord.[3] For instance, a chord's first Fourier component is approximately inversely related to the size of the minimal voice leading to any transposition of {0}; the second Fourier component is approximately inversely related to the size of the minimal voice leading to any transposition of either {0} or {0, 6}; the third component is approximately inversely related to the size of the minimal voice leading to any transposition of either {0}, {0, 4}, or {0, 4, 8}, and so on. Interestingly, however, we can see this connection clearly only when we model chords as multisets in continuous pitch-class space, following the approach of Callender, Quinn, and Tymoczko (2008). (This in fact may be one reason why previous theorists did not notice the relationship.) When we do adopt this perspective, we see that there is a deep relationship between two seemingly very different conceptions of set-class similarity, one grounded in voice leading, the other in interval content. Furthermore, this realization allows us to generalize some of the features of Quinn's approach, using related methods that transcend some of the limitations of the Fourier transform proper.

I. Voice leading and set-class similarity

Let me begin by describing the voice-leading approach to set-class similarity (or inverse distance), reviewing along the way some basic definitions. Much of what follows is drawn from (or implicit in) earlier essays, including Tymoczko 2006 and 2008 and Callender, Quinn, and Tymoczko 2008; readers who want to explore these ideas further are hereby referred to these more in-depth discussions.

We can label pitch classes using real numbers (not just integers) in the range [0, 12), with C as 0.[4] Here the octave has size 12, and familiar twelve-tone equal-tempered semitones have size 1. This system provides labels for every conceivable pitch class and does not limit us to any particular scale; thus, the number 4.5 refers to "E quarter-tone sharp," halfway between the twelve-tone equal-tempered pitch classes E and F.

A *voice leading between pitch-class sets* corresponds to a phrase like "the C major triad moves to E major by moving C down to B, holding E fixed, and shifting G up by semitone to G♯." We can notate this more efficiently by writing

---

3 By "perfectly even *n*-note chord" I mean the chord that exactly divides the octave into *n* equally sized pieces, not necessarily lying in any familiar scale. For example, the perfectly even eight-note chord is {0, 1.5, 3, 4.5, 6, 7.5, 9, 10.5}.

4 The notation [*x*, *y*) indicates a range that includes the lower bound *x* but not the upper bound *y*. Similarly, (*x*, *y*) includes neither upper nor lower bounds, while [*x*, *y*] includes both.

(C, E, G) $\xrightarrow{-1, 0, 1}$ (B, E, G♯), indicating that C moves to B by one descending semitone, E moves to E by zero semitones, and G moves to G♯ by one ascending semitone. The order in which voices are listed is not important; thus, (C, E, G) $\xrightarrow{-1, 0, 1}$ (B, E, G♯) is the same as (E, G, C) $\xrightarrow{0, 1, -1}$ (E, G♯, B). The numbers above the arrows represent paths in pitch-class space, or directed distances such as "up two semitones," "down seven semitones," "up thirteen semitones," and so on. When the paths all lie in the range (–6, 6] I eliminate them; thus, a notation like (C, E, G) → (B, E, G♯) indicates that each voice moves to its destination along the shortest possible route, with the arbitrary convention being that tritones ascend. Formally, voice leadings between pitch-class sets can be modeled as multisets of ordered pairs, in which the first element is a pitch class and the second a real number representing a path in pitch-class space.

Voice leadings are *bijective* when they associate each element of one chord with precisely one element of the other. However, it matters whether we represent chords as sets (containing no duplications) or multisets (which may contain multiple copies of pitch classes). For example, the voice leading (C, C, E, G) → (A, C, F, F) is simultaneously a nonbijective voice leading between the sets {C, E, G} and {F, A, C} and also a bijective voice leading between the multisets {C, C, E, G} and {F, F, A, C}. For the purposes of this article, it is convenient to represent chords as multisets and to consider only bijective voice leadings between them. However, in other contexts, it can be useful to consider sets and nonbijective voice leadings.[5] It turns out to be a nontrivial task to devise an algorithm for measuring set-class similarity when nonbijective voice leadings are permitted. Fortunately, this complication is irrelevant here.

We measure the *size* of a voice leading using some function of (or partial order on) the nondirected distances moved by the individual voices. (These are the absolute values of the numbers above the arrows in the voice leading.) In principle, there are many different measures of voice-leading size but no compelling reason to choose one over another (Tymoczko 2006; Hall and Tymoczko 2007). In this article, however, it is convenient to use the Euclidean metric, according to which the size of a collection of real numbers $x_1, x_2, \ldots, x_n$ is

$$\sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}.$$

The reasons for this choice are that the Euclidean metric (1) provides a reasonable approximation to a range of voice-leading measures (Hall and Tymoczko 2007), (2) is computationally tractable, and (3) is particularly well suited to the task of investigating the Fourier transform. The latter two points are clarified shortly.

We can define the *distance between two set-classes* as the size of the minimal voice leading between any of their transpositions or inversions. The term *any*

---

5 For example, one might consider the distance between C and E major seventh chords to be determined by the nonbijective voice leading (C, E, E, G, B) → (B, D♯, E, G♯, B), which is in fact smaller than the smallest four-voice voice leading between them. See Callender, Quinn, and Tymoczko 2008, supplementary section 7.

here means "any of their forms in *continuous pitch-class space*"; thus, when measuring distances between set-classes we cannot necessarily confine ourselves within any particular scale. For example, according to the Euclidean metric, the distance between the perfect fourth and major third is given not by the voice leading (C, F) → (D♭, F), with size 1, but by (C, F) → (C♯, E♯) (or C "quarter-tone sharp," E "quarter-tone sharp") with size

$$\sqrt{\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2} = 1/\sqrt{2} \cong 0.707.$$

Though this may initially seem counterintuitive, it has on reflection a certain logic: if we are really interested in *intrinsic* relations between set-classes, then there is no reason to think that we can limit our attention to those that happen to appear in any one scale.

Now the Euclidean metric is particularly convenient for the following reason: if we are looking for the minimal voice leading between any two transpositions of any two chords, we need only consider those whose pitch classes sum to the same value modulo 12. (This in turn follows from basic facts of Cartesian geometry.)[6] For example, suppose we are trying to find the minimal Euclidean voice leading from the C augmented triad to any diminished triad. The pitch classes {C, E, G♯} are represented by the numbers {0, 4, 8}, which sum to $0 + 4 + 8 = 12 \equiv 0$ (mod 12). To find the nearest diminished triad, we need only consider those whose pitch classes sum to 0 (mod 12): {0, 3, 9}, {1, 4, 7}, and {5, 8, 11}. Observe that there are three, all related by major-third transposition. In general, we can always transpose an *n*-note chord by $12/n$ semitones without changing its sum (mod 12), and we can repeat this procedure *n* times before the initial chord reappears; thus, there will in general be *n* different transpositions of each *n*-note chord summing to the same number.[7] Note also that when searching for minimal voice leadings, we will frequently need to consider fractional pitch classes; for example, to find the nearest minor triad to {0, 4, 7}, we need to look at those summing to 11. These are

$$\{0\tfrac{1}{3}, 3\tfrac{1}{3}, 7\tfrac{1}{3}\}, \{4\tfrac{1}{3}, 7\tfrac{1}{3}, 11\tfrac{1}{3}\}, \{3\tfrac{1}{3}, 8\tfrac{1}{3}, 11\tfrac{1}{3}\}$$

(or, in other words, the familiar C minor, E minor, and A♭ minor triads, transposed up by one-third of a semitone). These chords, of course, do not reside on the ordinary piano keyboard.

Finally, suppose that we have two chords $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ with each chord's pitch classes listed in ascending numerical order (when

---

6 An ordered set can be modeled as a point in $\mathbb{R}^n$. Transposition corresponds to motion along the "unit diagonal" that contains both the origin and (1, 1, . . . , 1). Transpositional set-classes can thus be represented by *lines* parallel to the unit diagonal. The shortest vector between any two of these lines will (according to the Euclidean metric) be perpendicular to both. This means that the vector's dot product with

(1, 1, . . . , 1) will be equal to zero, which in turn implies that the sum of its components is zero. Hence, the coordinates of its endpoints sum to the same value.

7 The qualification "in general" is needed because of symmetrical chords: when we transpose {0, 4, 8} by four semitones, we get the same chord again.

considered as real numbers). To find the minimal voice leading between them, we need to consider n different circular permutations

$$(x_1, x_2, \ldots, x_n) \rightarrow (y_1, y_2, \ldots, y_n),$$

$$(x_1, x_2, \ldots, x_n) \rightarrow (y_2, y_3, \ldots, y_n, y_1),$$

$$.$$
$$.$$
$$.$$

$$(x_1, x_2, \ldots, x_n) \rightarrow (y_n, y_1, y_2, \ldots, y_{n-1}).$$

(These voice leadings have been described as the n "interscalar transpositions" between the chords; that they are the only possibilities follows from the fact that voice crossings always increase the Euclidean size of a voice leading [Tymoczko 2006].) For example, to identify the minimal voice leading between {0, 3, 9} and {3, 10, 11}, we need to consider the three voice leadings

$$(0, 3, 9) \xrightarrow{3, -5, 2} (3, 10, 11),$$

$$(0, 3, 9) \xrightarrow{-2, -4, 6} (10, 11, 3),$$

$$(0, 3, 9) \xrightarrow{-1, 0, 1} (11, 3, 10).$$

Clearly, the third is the smallest, with a total size of $\sqrt{2}$. It may again seem strange that we have to consider all these possibilities: roughly speaking, the reason is that there is no way to determine the destination of any particular pitch class without calculating the size of each and every one of these voice leadings. In particular, we have no assurance that a maximally efficient voice leading always associates a pitch class in one chord with its nearest neighbor in the other.

Putting it all together, then, we can use the following procedure to find the minimal Euclidean voice leading between two n-note multiset-classes A and B:

(1)   Choose a representative of A and calculate the sum of its pitch classes.
(2)   Find the n transpositions of B that sum to this same value.
(3)   For each of these, calculate the (Euclidean) size of the n "interscalar transpositions" described in the previous paragraph.
(4)   Repeat steps 2 and 3 for the inversion of B.
(5)   The minimum of these $2n^2$ numbers is the Euclidean distance between the multiset-classes.

Though it would be somewhat laborious to follow this algorithm by hand, it is easy to program a computer to do it. The result is a single number representing the Euclidean distance between set-classes. Equivalently, this number can be taken to represent the voice-leading distance from any particular set

(e.g., the C augmented chord) to the *nearest* transposition or inversion belonging to some other set-class (e.g., any diminished triad in continuous space).

II. Voice-leading distances and Fourier magnitudes

To explore the connection between voice leading and the Fourier transform, it is useful to begin with the "set-class spaces" described by Callender, Quinn, and myself. These are n-dimensional geometrical spaces containing all the multiset-classes of size n, where distances are as described in the preceding section.[8] Figure 1 shows the location, in three-note set-class space, of the multiset-classes that can be constructed using the pitches of some perfectly even n-note chord, for n ranging from 1 to 6.[9] (Terminological note: I refer to these as the "doubled subsets of the perfectly even n-note set-class.")[10] Associated to each graph is one of the six Fourier components. For any three-note set-class, the magnitude of its nth Fourier component is a decreasing function of the distance to the nearest of these marked points; for instance, the magnitude of the third Fourier component (FC_3) decreases the farther one is from the nearest of $\{0, 0, 0\}$, $\{0, 0, 4\}$, and $\{0, 4, 8\}$. Thus, set-classes in the shaded region of Figure 2 will tend to have a relatively large FC_3, while those in the unshaded region will have a smaller FC_3.

　　Figure 3 presents three-dimensional graphs in which the x,y-plane represents triangular set-class space, as in Figures 1 and 2, and where the z-axis represents the magnitude of the relevant Fourier component.[11] The graphs show a series of peaks precisely at the doubled subsets of the perfectly even n-note set-class, with valleys at the points most distant from these peaks. It is clear from the graphs that there is a decreasing relationship between height (nth Fourier magnitude) and distance to the nearest peak (doubled subset of the perfectly even n-note set-class). Furthermore, the contour lines, showing set-classes of equal Fourier magnitude, are roughly circular. This means that the relevant measure of voice-leading size is the Euclidean metric, as this is the metric for which a circle's points are equidistant from the center.[12] This is quite fortunate, since Euclidean distance is also particularly easy to work with, for the reasons discussed above.

8 See Callender 2004; Tymoczko 2006; Callender, Quinn, and Tymoczko 2008. Mathematically, these spaces are the quotients of tori both by central inversion and by cyclical permutations of their barycentric coordinates.

9 Cliff Callender, in a personal communication, points out that the marked points in Figure 1 depict portions of a regular lattice and that they differ only by a multiplicative factor.

10 These are not "submultisets" since they may introduce additional duplications: {0, 0, 4} is not a submultiset of {0, 4, 8}, since the former contains two copies of the "0" while the latter contains only one. However, the latter chord can be constructed by introducing doublings into a subset of the perfectly even chord, hence the term "doubled subsets."

11 Thanks to Cliff Callender for programming assistance. A very similar graph appears in Callender 2007, which explores the Fourier transform in continuous space.

12 There are many reasonable ways to measure voice leading, as emphasized in both Tymoczko 2006 and Hall and Tymoczko 2007. Each produces a different set of points equidistant from a given location: for the "taxicab" metric, this set is a diamond; for the Euclidean metric, a circle; and for the "largest distance" metric, a square. See Hall and Tymoczko 2007 for further discussion.

FC$_1$, subsets of {0}

FC$_2$, subsets of {0, 6}

FC$_3$, subsets of {0, 4, 8}

FC$_4$, subsets of {0, 3, 6, 9}

FC$_5$, subsets of {0, 2.4, 4.8, 7.2, 9.6}

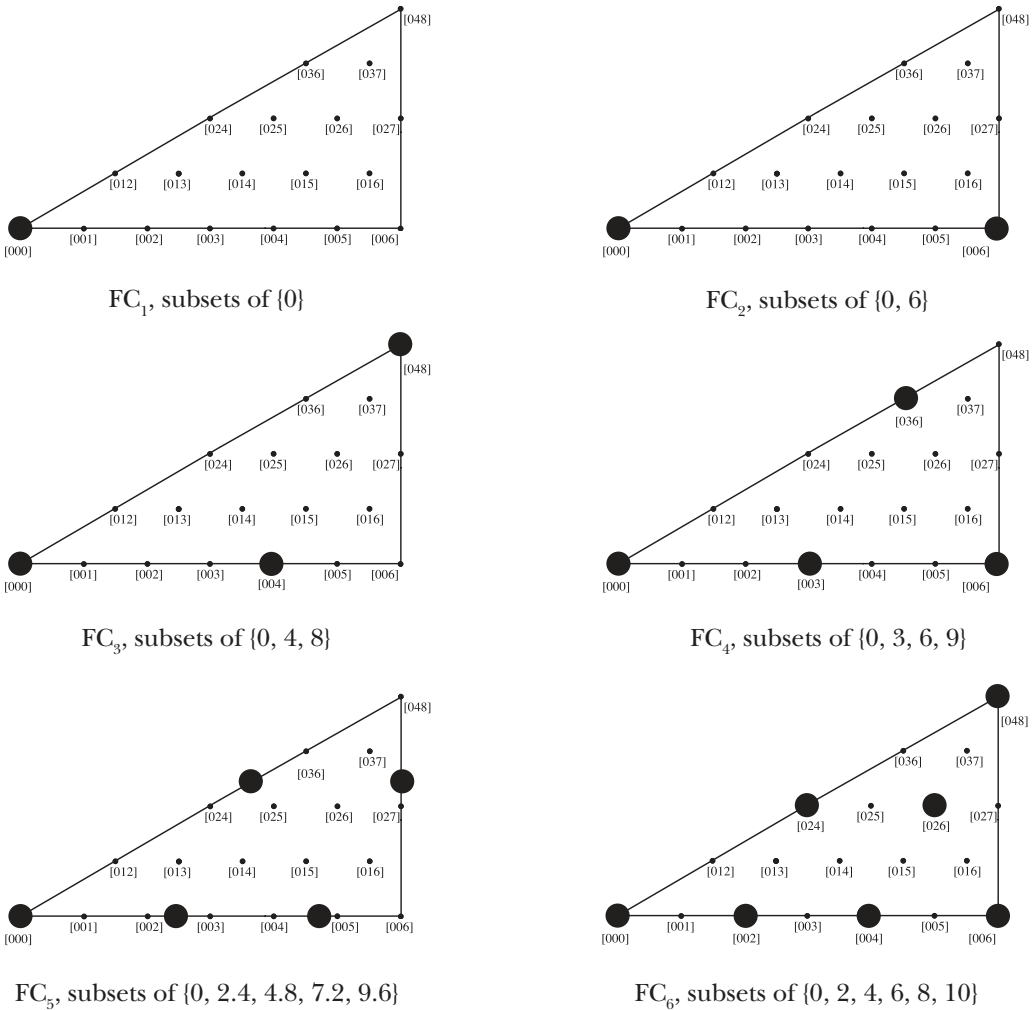FC$_6$, subsets of {0, 2, 4, 6, 8, 10}

Figure 1. The magnitude of a set-class's *n*th Fourier component is approximately inversely proportional to the distance to the nearest doubled subset of the perfectly even *n*-note set-class, shown here as dark spheres. Each graph corresponds to one Fourier component and one collection of doubled subsets.

Figure 3 provides intuitive evidence that there is a connection between the *n*th Fourier magnitude and distance to the nearest doubled subset of the perfectly even *n*-note set-class. The next task is to quantify and to explain this relationship. I begin with a brute-force calculation that relates the two quantities for every twelve-tone equal-tempered multiset-class. (Other temperaments are considered below.) The first part of the calculation is easy, requiring only
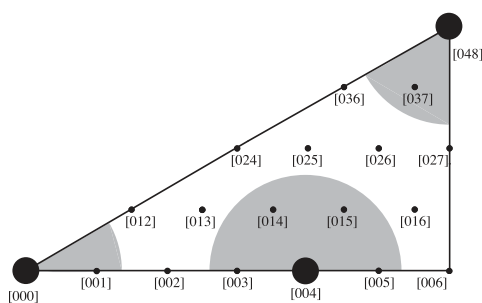
Figure 2. Set-classes in the shaded region will have a large third Fourier component, since they are near doubled subsets of {0, 4, 8}. Those in the unshaded region will have a smaller third Fourier component.

Equation 1, above (Figure 4). The second is a little more difficult: in principle, we need to repeat the algorithm in Section I for each doubled subset of the perfectly even *n*-note set-class.[13] However, Section III describes a shortcut that simplifies the calculation considerably.

Once we determine both the *n*th Fourier component and the minimal voice leading to the nearest doubled subset of the *n*-note set-class, we can plot these two numbers for every (twelve-tone equal-tempered) multiset-class of a given cardinality. Figure 5 shows, for trichordal multiset-classes, both the FC_3 magnitude and the size of the minimal voice leading to the nearest doubled subset of {0, 4, 8}. It is clear that there is a very nearly linear relation between these two quantities, illustrated by the gray line:

$$FC_3 = -1.38VL + 3.16. \tag{2}$$

Using this equation, one can estimate a trichord's third Fourier component (FC_3) on the basis of the minimal voice leading to the nearest doubled subset of any augmented triad (VL), and vice versa. The Pearson correlation coefficient is a standard statistical measure that quantifies the "degree of fit" between the points and the line. Here, the value –0.97 indicates that there is a very nearly linear relation between the values.[14]

Table 1 correlates voice-leading distances and Fourier components, for twelve-tone equal-tempered multiset-classes of other cardinalities. The values in the table are determined by carrying out the computations in Figure 4 for every equal-tempered multiset-class of size 2–10, and every Fourier component from 1 to 6. (Appendix 1S, which appears as supplemental material [online only] with this article at http://dx.doi.org/10.1215/00222909-2009-019, presents the raw data necessary to reconstruct Table 1.) The strong anticorrelations indicate that one variable predicts the other with a high degree of accuracy.

---

13 If we are considering a *k*-element set-class, we need to construct all of those doubled subsets with *k* elements. Thus, for the third Fourier component and three note chords, we need {0, 0, 0}, {0, 0, 4}, and {0, 4, 8}.

14 A correlation of –1 indicates a perfect decreasing linear relation; a correlation of +1, a perfect increasing linear relation; and a correlation of 0, no linear relationship at all.
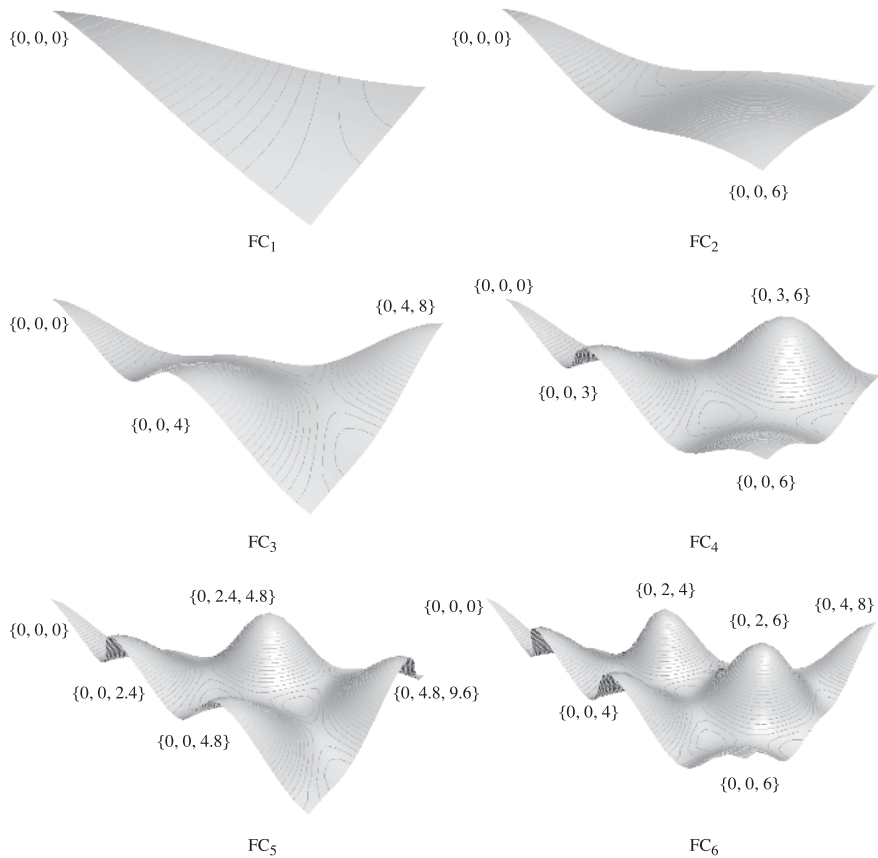
Figure 3. Here the x,y-plane represents triangular set-class space, while the z-axis represents Fourier magnitudes. The peaks are located at the doubled subsets of the perfectly even *n*-note set-classes.

III. Understanding the correlations

We now explore this relationship in a more rigorous way. It follows from Equation 1 that the *n*th Fourier component represents pitch classes as unit vectors in a "reduced" pitch-class space whose octave is only $12/n$ semitones large. (The factor $2\pi n/12$ maps pitch classes in the range $[0, 12/n)$ to the circumference of the unit circle; larger pitch-class numbers are reduced modulo $12/n$.)[15] Since all pitch classes p and $p + 12/n$ will be represented by identical vectors, we can move any note by $12/n$ semitones without changing the *n*th Fourier component (see Figure 5; see also Hoffman 2008).[16] This is illustrated geometrically in Figure 6. As long as pitch-class space is quantized finely

15 The reduced octave also appears in Cohn 1991.

16 Some equal temperaments will not contain both p and $p + 12/n$. (E.g., twelve-tone equal temperament does not contain p and $p + 2.4$.) However, we can always embed an equal temperament into a more finely grained equal temperament containing p and $p + 12/n$. For the purposes of conceptualizing the Fourier transform, it is often useful to work in this more finely quantized space, or in continuous unquantized space.

a) Calculating the third Fourier component (FC$_3$) of {0, 2, 5}.

  Step 1: assign the vector $(\cos 2\pi pn/12, \sin 2\pi pn/12)$ to each pitch class *p*.

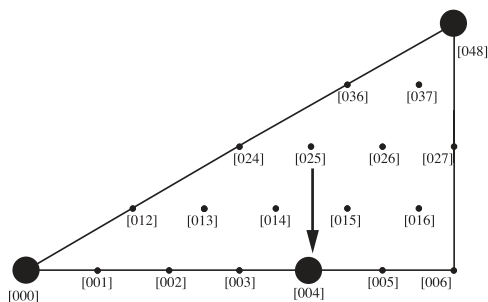 $0 \rightarrow (\cos(2\pi(0 \times 3)/12), \sin(2\pi(0 \times 3)/12)) = (1, 0)$

    $2 \rightarrow (\cos(2\pi(2 \times 3)/12), \sin(2\pi(2 \times 3)/12)) = (-1, 0)$

    $5 \rightarrow (\cos(2\pi(5 \times 3)/12), \sin(2\pi(5 \times 3)/12)) = (0, 1)$

  Step 2: add these vectors: $(1, 0) + (-1, 0) + (0, 1) = (0, 1)$

  Step 3: determine the length of this sum: $\|(0, 1)\| = \sqrt{0^2 + 1^2} = 1$

b) Determining the distance to the nearest doubled subset of {0, 4, 8}.

Following the procedure in Section I, we learn that $(0, 2, 5) \rightarrow (1, 1, 5)$ is the minimal voice leading, moving two voices by one semitone each, and with a Euclidean size of $\sqrt{1^2 + 1^2} \cong 1.41$

Figure 4. Calculating the size of the third Fourier component of {0, 2, 5} and the minimal voice leading from {0, 2, 5} to any doubled subset of {0, 4, 8}
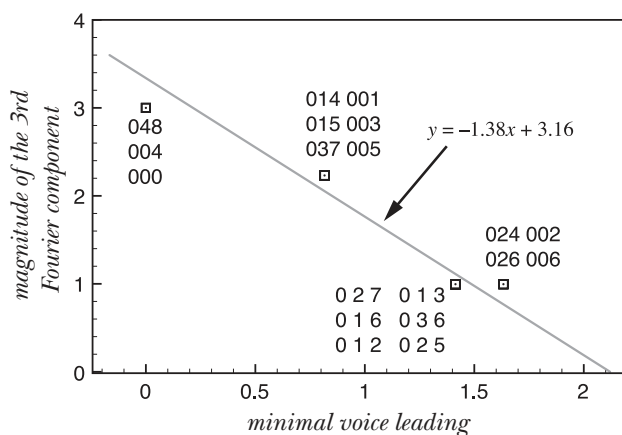

Figure 5. For trichords, the equation FC$_3$ = –1.38VL + 3.16 relates the third Fourier component to the Euclidean size of the minimal voice leading to the nearest doubled subset of {0, 4, 8}.

enough, moving any note by chromatic step will cause only a minimal change to the Fourier components.[17] To determine the magnitude of a chord's *n*th Fourier component, we add the vectors representing the all notes in the chord and calculate the length of the result.

---

17 When pitch-class space is not finely quantized, this will not always be the case. For instance, consider the fifth Fourier component in twelve-tone equal temperament. The pitch classes 0, 2.4, 4.8, 7.2, and 9.6 are assigned the same vectors in the reduced pitch-class space of length 2.4. Moving 0.2 of a semitone on the reduced circle leads to a point

Table 1. Correlations between voice-leading distances and Fourier magnitudes

| Chord Size | Fourier Component | | | | | |
	FC_1	FC_2	FC_3	FC_4	FC_5	FC_6
Dyads	−0.97	−0.96	−0.97	−1	−0.97	−1*
Trichords	−0.98	−0.97	−0.97	−0.98	−0.98	−1*
Tetrachords	−0.96	−0.96	−0.95	−0.98	−0.96	−1*
Pentachords	−0.96	−0.96	−0.95	−0.98	−0.96	−1*
Hexachords	−0.96	−0.96	−0.95	−0.96	−0.96	−1*
Septachords	−0.96	−0.96	−0.96	−0.97	−0.96	−1*
Octachords	−0.96	−0.96	−0.95	−0.98	−0.96	−1*
Nonachords	−0.96	−0.96	−0.96	−0.98	−0.96	−1*
Decachords	−0.96	−0.96	−0.96	−0.98	−0.96	−1*

*Voice leading calculated using L^1 (taxicab) distance rather than L^2 (Euclidean)

By contrast, when thinking about voice leading, we represent pitch classes not as vectors but rather *points on the pitch-class circle*. Typically, we take the circumference of the circle to be one octave. Suppose, however, that we would like to find the minimal voice leading from some chord to the nearest doubled subset of a perfectly even *n*-note chord. Figure 7 shows that this problem is closely related to the problem of finding a voice leading from the image of this pitch-class set, in a reduced pitch-class circle of circumference
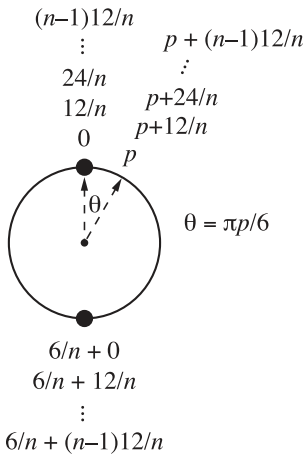

Figure 6. The *n*th component of the Fourier transform imposes a smaller periodicity on the pitch-class circle, representing pitches *p* and *p* + 12*i*/*n* by the same vector, for all integers *i*. To determine the chord's *n*th Fourier component, we add all the vectors corresponding to the notes in the chord.

representing pitch classes 0.2, 2.6, 5, 7.4, and 9.8. Thus the perfect fourth appears to be smaller than the semitone—indeed, it is the smallest twelve-tone equal-tempered interval. (Hoffman 2008 exploits this fact to draw voice-leading lattices in which notes move by perfect fifth.) When we quantize more finely, however, motion by 0.2 of a semitone is seen to be just as small as motion by five semitones, and motion by 0.1 of a semitone is smaller still.
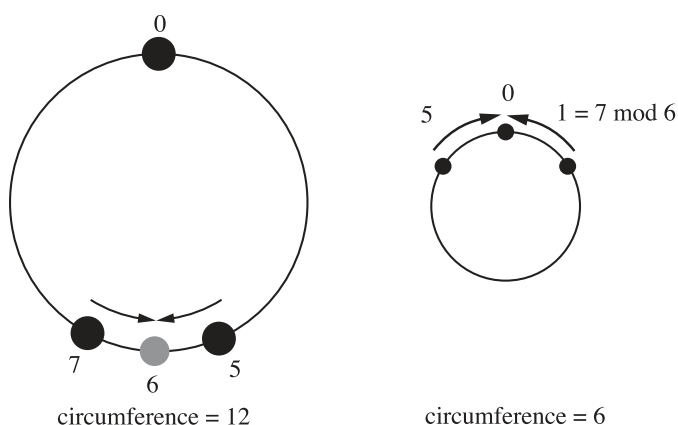
circumference = 12

circumference = 6

Figure 7. In searching for the minimal voice leading from any chord to the nearest doubled subset of any transposition of the perfectly even *n*-note chord, it is sufficient to represent the initial chord on a reduced pitch-class circle of size 12/*n*. The figure on the left represents the minimal voice leading from {0, 5, 7} to any subset of {0, 6}, which is (0, 5, 7) → (0, 6, 6). The figure on the right shows that this corresponds to the voice leading (0, 1, 5) → (0, 0, 0) in the reduced pitch-class space of size 6.

$12/n$, to some transposition of the unison $\{0, \ldots, 0\}$: any voice leading from a set S to a doubled subset of the perfectly even n-note chord determines a unique voice leading from the image of S to a unison in the reduced pitch-class space.[18] Thus, we need only look for voice leadings to doubled unisons in the reduced pitch-class circle of length $12/n$. This allows us to improve our algorithm for identifying minimal voice leadings to the nearest doubled subset of a perfectly even n-note chord.[19]

The reduced pitch-class circle of length $12/n$ therefore arises both in determining the nth Fourier component and in identifying the minimal voice leading to the nearest doubled subset of any perfectly even n-note chord. The next task is to understand the relationship quantitatively. Figure 8 shows that a collection of vectors will yield the largest sum when they are all pointing in the same direction or, in other words, when the chord they represent is a doubled subset of the perfectly even n-note chord. The vectors will yield the smallest sum when they point in directions that are evenly distributed around the reduced pitch-class circle and hence cancel each other out.

---

18 In more mathematical terms: any voice leading in the larger pitch-class space, from a set S to any subset of the perfectly even n-note chord, will *project* to an equally sized voice leading in the reduced pitch-class space, from the *image* of set S to a unison; conversely, any voice leading in the reduced space, from any set to a unison, can be *lifted* to a collection of equally sized voice leadings in the larger space. These voice leadings link the *preimage* of the set S to a doubled subset of some perfectly even n-note chord.

19 We begin by representing the chord modulo 12/*n*; we then consider the unisons whose pitch classes sum to the same value (mod 12/*n*) as those in the original chord. Thus, if $\{x_1, x_2, \ldots, x_m\}$ is our chord (mod 12/*n*), with $x_1 + x_2 + \ldots + x_m \equiv s$ (mod 12/*n*), we need only consider voice leadings to the unison (s/m, s/m, $\ldots$, s/m) and its transpositions by 12/*nm* semitones.

Conversely, the size of the minimal voice leading to the nearest unison will be *zero* when the vectors point in the same direction and will be maximally large when the vectors are evenly distributed around the circle. Thus, there should be a decreasing relation between the magnitude of the nth Fourier component and the minimal voice leading to the nearest subset of the perfectly even n-note chord.
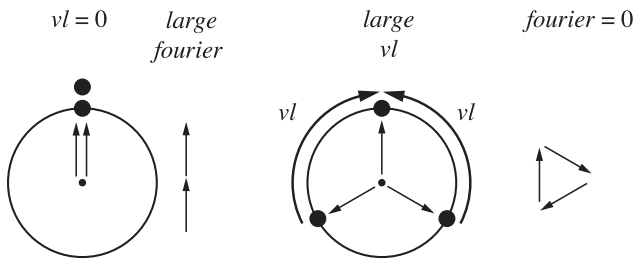

Figure 8. (Left) Doubled subsets of a perfectly even *n*-note chord will have a large *n*th Fourier component, since they will be represented by vectors pointing in the same direction. No voice leading is necessary to transform these chords into doubled subsets of the perfectly even *n*-note chord. (Right) Chords whose vectors are evenly distributed around the reduced pitch-class circle will have an *n*th Fourier component of zero, since their vectors cancel out. It takes a large voice leading to move these chords to a unison in the reduced pitch-class circle.

Importantly, however, the relation is only approximate. This is because the Fourier transform represents pitch classes as *vectors* while the voice-leading model represents them as *points on the circle*. Mathematically these are very different, and there is no reason to expect the calculations to correspond precisely. In fact, there exist chords that have the same nth Fourier component but are not equidistant from the nearest doubled subset. Figure 9, for example, shows two collections of vectors summing to zero: if we assemble a chord by choosing one of the pitch classes at the head of each vector, it will have a zero FC_3 component; however, chords constructed using the circle on the right will be slightly closer to the nearest doubled subset of the nearest augmented triad. (See the tables in the online supplement to this article.) Thus, the best we can hope for is an approximate correspondence between Fourier magnitudes and voice-leading distances.

Indeed, this can be seen directly from Figure 3: the "hills" are not perfect cones, which implies that the relation between voice-leading distance and Fourier magnitude is not perfectly linear. Furthermore, though it is less apparent from the graph, the contour lines are not exactly circles, which means that there is no function that will calculate Fourier magnitudes solely on the basis of Euclidean voice-leading distances. This is shown more clearly in Figure 10, which plots the magnitude of the first Fourier component against voice-leading distance from the nearest tripled unison ($\{0, 0, 0\}$), for all trichordal multiset-classes in 192-tone equal temperament. (This graph is the analogue
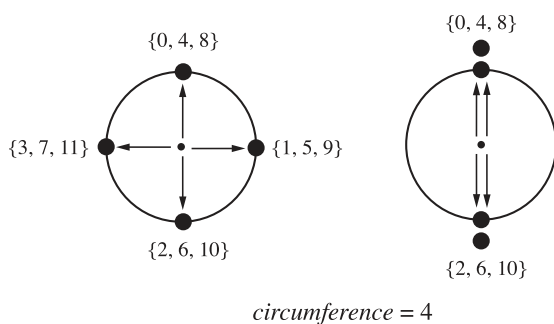
circumference $= 4$

Figure 9. For each circle, one can assemble a number of different multisets by choosing one pitch class at the head of each arrow. All of these will have a vanishing third Fourier component. However, those produced by the rightmost circle will have a slightly smaller voice leading to the nearest subset of the nearest augmented triad.

of Figure 5 for this very finely quantized chromatic universe.) For chords close to the triple unison, there is basically a one-to-one correspondence between Fourier magnitude and voice-leading distance, as can be seen from the fact that the upper-left portion of the graph is very thin. (Note that the slight curvature indicates that the relationship is not quite linear.) The "bulge" on the lower right shows that the relation becomes more approximate with increasing distance: here, multiset-classes can have a range of Fourier magnitudes, even if they are equidistant from the triple unison. The graph tapers again for chords maximally distant from {0, 0, 0}, indicating that the relation between voice leading and Fourier magnitudes becomes more precise at large distances. Figure 10 thus clearly shows both that voice-leading distance is a reasonable predictor of the Fourier magnitude and that the relationship is necessarily somewhat approximate. We cannot perfect our predictions simply by using another familiar measure of voice leading, or even a simple function thereof: since there is essentially a one-to-one relationship near the triple unison, any equation relating Fourier magnitudes to these voice-leading distances must reduce to the Euclidean metric at short range. However, because of the bulge in Figure 10, we know that at larger distances anything resembling the Euclidean metric will provide only an approximate predictor of the magnitude of the first Fourier component.

Figure 11 contains analogous graphs relating Fourier magnitudes to voice-leading distances for tetrachordal, pentachordal, and hexachordal multiset-classes in 48-tone equal temperament. The graphs are all reasonably similar in shape. Unlike Figure 10, they do not "taper" at the point of maximal distance from the perfect unison.[20] The graphs are increasingly dense for

---

20 This is because there is only one way (within transposition) to arrange three unit vectors so that they sum to zero, whereas there are several ways of doing it for four or more vectors. Note that the graphs for a k-note chord have a pronounced inflection point at Fourier magnitude $k - 2$. This may reflect the fact that there are a large number of ways to combine $k - 2$ vectors pointing in the same direction with two other vectors pointing opposite one another.
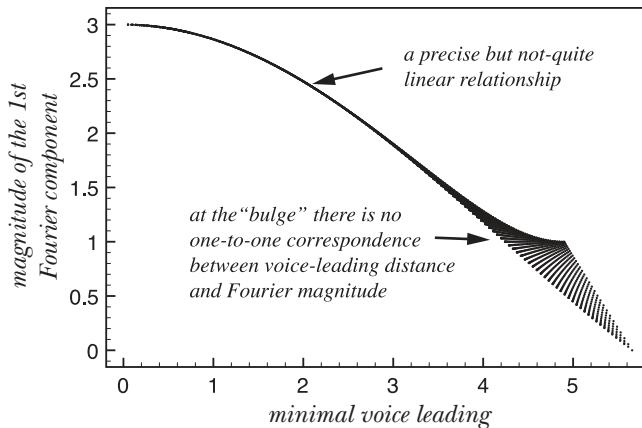
Figure 10. Fourier magnitudes and voice-leading distance for trichords in 192-tone equal temperament. The correlation between the two values is –0.99.

larger chords, reflecting the fact that the number of multiset-classes grows very quickly with cardinality. (Indeed, there are about a quarter-million hexachordal multiset-classes in 48-tone equal temperament, and even more for higher cardinalities—which is why it is difficult to produce analogous graphs for larger multisets.) Table 2 calculates the correlation between voice leading and Fourier magnitudes for three- to six-note chords in 48-tone equal temperament. The strong anticorrelations show that relationship continues to hold in very finely quantized pitch-class space. (In fact, 48-tone equal temperament is dense enough that these values approximate those for unquantized, continuous pitch-class space.)[21] Furthermore, in continuous space, the graphs of all the Fourier components will be essentially identical, since in each case vectors can point in any direction on the relevant reduced pitch-class circle. Thus, the graphs in Figures 10 and 11, as well as the correlations in Table 2, can be taken to represent not just the first Fourier component but the other components, as well.

In my view, we should not be disappointed that there is only an approximate relation between voice-leading distance and Fourier magnitude. Both the Fourier transform and the Euclidean voice-leading metric are very precise tools for modeling inherently vague musical intuitions, and we should not become too invested in their fine quantitative structure; indeed, there is little reason to think that very small differences in *either* Fourier magnitude or Euclidean voice-leading distance correspond to anything psychologically real for composers or listeners. What is more interesting, to my mind, is that both the Fourier transform and voice leading provide similar, and intuitively
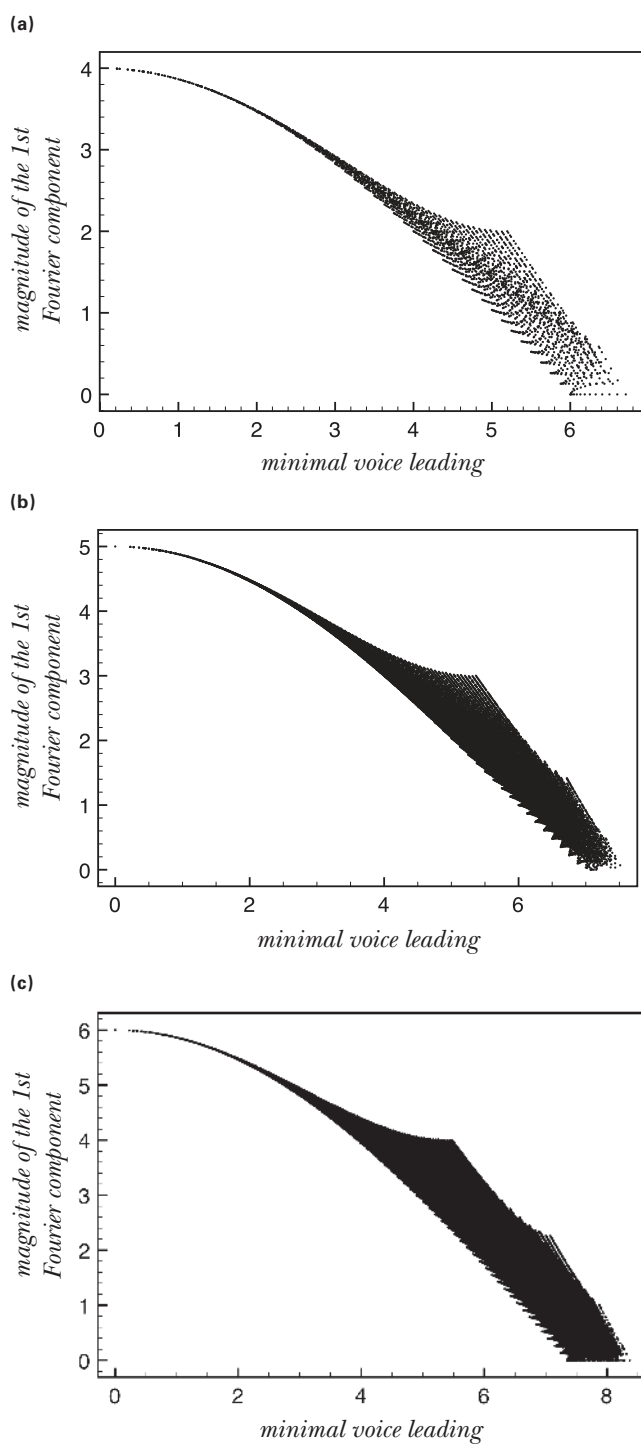
(a)

(b)

(c)

Figure 11. Fourier magnitudes and voice-leading distance for tetrachords (a), pentachords (b), and hexachords (c) in 48-tone equal temperament

Table 2. Correlations between voice-leading distances and Fourier magnitudes in 48-tone equal temperament. These numbers are approximately valid for any Fourier component in continuous pitch-class space.

Chord Size	FC_1
Trichords	−0.99
Tetrachords	−0.97
Pentachords	−0.97
Hexachords	−0.96

plausible, ways of modeling the sense that chords can be very "major thirdy" (or "whole-tonish") without being *exactly* so. Here it is important that there is a particularly strong resemblance for chords very close to doubled subsets of perfectly even n-note chords. Thus, the two models will agree about which chords are especially "fifthy," "whole-tony," and so forth—even if they disagree somewhat about chords that are only mildly so.

Readers will have noticed that there is one circumstance in which Fourier facts precisely mirror voice-leading facts: for twelve-tone equal-tempered chords, the FC_6 magnitude records the absolute value of the difference between the number of its notes in one whole-tone scale and the number of its notes in the other (Figure 12). (Mathematically, this is a *scalar* rather than vector quantity.) One can obviously voice lead such chords to a doubled subset of a whole-tone scale simply by moving all of the notes in the less populous whole-tone set by semitone. It follows that the FC_6 values will be perfectly anticorrelated with the voice-leading distances obtained using the "taxicab" (rather than Euclidean) metric. In fact, for k-note chords, the equation $FC_6 = k - 2VL$ exactly determines the sixth Fourier component on the basis of voice leading, where VL is the taxicab distance to the nearest doubled subset of any whole-tone scale.
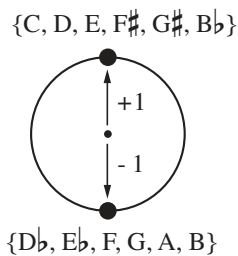
{C, D, E, F♯, G♯, B♭}

{D♭, E♭, F, G, A, B}

Figure 12. The sixth Fourier component assigns the value +1 for notes in one whole-tone scale and −1 for those in the other. The absolute value of the result represents the difference between the number of notes in the more and less populous whole-tone scales. This Fourier component is perfectly anticorrelated with the size of the voice leading to the nearest doubled subset of the nearest whole-tone scale—as long as we measure voice-leading distance using the "taxicab" metric.

IV. Discussion

Let's return to the thought that the Fourier transform models the way chords can be more or less saturated with particular intervals—that is, more or less chromatic, whole-tonish, or perfect fifthy. On one level, this seems accurate: chords such as {0, 2, 4} and {0, 0, 2} have a high sixth Fourier component, and they are indeed saturated with major seconds. But when we think more carefully, we notice that the simple statement is not quite right: {0, 4, 8} also has a very large sixth Fourier component, even though it contains no major seconds at all! Furthermore, the Fourier components of the tripled unison {0, 0, 0} are all maximally large, even though the multiset contains no nonzero intervals. (By the continuity of the Fourier transform, something similar is true of such chords as {0, e, 2e} for very small e.) Even the interpretation of the fifth Fourier component, as representing the "perfect fifthiness," needs to be qualified: in very finely quantized equal temperaments, chords such as {0, 2.4, 4.8, 7.2, 9.6}, which have no perfect fifths, have a larger fifth Fourier component than the pentatonic scale.

These examples suggest that we might sometimes want to depart from Fourier analysis in favor of an approach based on voice leading. The Fourier transform requires us to measure a chord's "harmonic quality" in terms of its distance from *all* the doubled subsets of the perfectly even set-classes. But we might sometimes wish to choose a different set of harmonic prototypes. For instance, Figure 13 uses distance from the augmented triad to measure trichordal set-classes' "augmentedness." Unlike Fourier analysis, this purely voice-leading–based method does not consider the triple unison or doubled major third to be particularly "augmented-like"; hence, set-classes like {0, 1, 4} do not score particularly highly on this index of "augmentedness." Similarly, we might sometimes wish to use a justly tuned diatonic scale as a harmonic prototype, rather than accepting the fifth Fourier component as a proxy for "diatonicness." (Suppose we are investigating the acoustic purity of the intervals in various temperaments' best diatonic scales; here, voice leading will produce much better results than the Fourier transform.) An approach based on voice leading leaves us free to choose the harmonic prototypes we want, rather than meekly accepting those the Fourier transform imposes on us.[22]

One way to put the point is that the Fourier transform is something of a black box: we put a chord in, and get some numbers out. (In fact, it can be quite hard to provide an intuitive characterization of what the Fourier transform actually does—particularly if one makes no reference to voice leading.) It is interesting that Quinn developed his Fourier-based technique under the influence of an avowedly "Platonist" conception of music theory, according to which "chord quality" is a fundamentally objective feature

---

22 This is not to suggest that we should abandon Fourier analysis in favor of approaches based on voice leading: Fourier analysis provides an indispensable tool for investigating questions involving the interval vector, as suggested in Lewin 1959 and 2001.
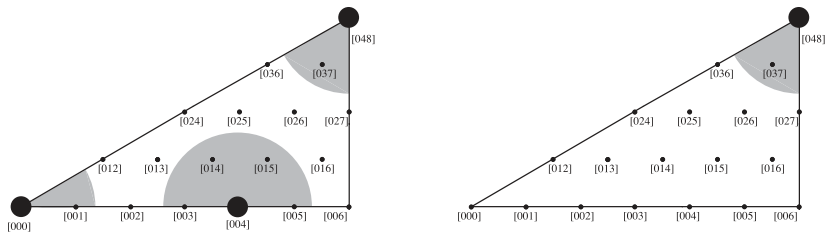
Figure 13. The mathematics of the Fourier transform requires that we conceive of "chord quality" in terms of the distance to *any doubled subset* of some perfectly even set-class (left). If we use voice leading, however, we can choose our harmonic prototypes freely. Thus, we can use voice leading to model a set-class's "augmentedness" in terms of its distance from the augmented triad (right), but not the tripled unison {0, 0, 0} or the doubled major third {0, 0, 4}.

that is (as it were) "out there in the world." By contrast, the voice-leading approach is consonant with a more relativist conception according to which we *choose* the musical properties that are important to us. A Platonist (e.g., the youthful Quinn) might well be attracted to the "black box" quality of the Fourier transform *precisely because of its inflexibility*—which could be taken to suggest an idealized world of unalterable musical relationships. And conversely, the very flexibility of the voice-leading approach might signal a (disturbing to some, attractive to others) role for arbitrary human preferences and choice.

Beyond measuring the intervallic saturation of single set-classes, we can of course use the Fourier transform to measure similarity *between* set-classes: from this point of view, set-classes are similar when their six Fourier magnitudes are all similar. At first blush, this strategy seems to contrast dramatically with the voice-leading approach: certainly, Fourier analysis uses very different mathematics, and produces results—such as the identity of Z-related chords—that can be difficult to interpret in contrapuntal terms.[23] We have seen, however, that there is a close relationship between the two techniques: at the most fundamental level, each individual Fourier component measures something like a voice-leading distance. Thus what is distinctive about the Fourier approach to chord similarity is not the conception of distance per se, but rather the role of "harmonic prototypes": the Fourier transform measures the similarity of set-classes not by their distance *from one another* but by their respective distances from the nearest doubled subsets of the perfectly even *n*-note chords. This is why Z-related chords are judged to be identical, even while being far apart in the set-class spaces such as Figure 1.[24]

From my point of view, the most interesting result is that a single conception of musical distance—voice-leading distance—turns out to underlie both approaches. It is, I think, quite surprising that voice leading should play any

23 See Quinn 2006 and 2007. Quinn's approach is inspired by earlier writers who emphasize shared subset content and the interval vector. See Quinn 2001 for more discussion.

24 Even from a voice-leading perspective, two Z-related chords will be approximately equidistant from the nearest doubled subsets of perfectly even *n*-note chords.

role whatsoever in the Fourier transform, with its vectors, trigonometric functions, and sensitivity to chords' interval content. That we can reinterpret its results *contrapuntally* says something about the power of an approach that puts voice leading front and center. In fact, one might even take it to suggest that Quinn's early Platonism was not entirely misplaced: perhaps Quinn was right to think that there is a realm of objective musical relationships that influence us even when we are not directly aware of them. (Certainly, not everything in music theory can be a matter of arbitrary personal preference!) If so, then I would argue that voice leading—rather than the Fourier transform—has the best claim to Platonic primacy. Perhaps it is spaces like Figure 1 that offer the best glimpse of the entities casting shadows on the walls of our musical cave.

Appendix

The raw data from which Table 1 was constructed appears as supplemental material (online only) with this article at http://dx.doi.org/10.1215/00222909-2009-019. Appendix 1S shows the Fourier magnitudes and corresponding minimal voice leadings for all twelve-tone equal-tempered multiset-classes. Appendix 2S contains the data for twelve-tone equal-tempered set-classes.

An individual table is provided for set-classes and multiset-classes of each cardinality: the first column identifies the (multi)set-class; the second shows the first Fourier magnitude; the third, the size of the minimal voice leading to the nearest doubled unison; the fourth, the second Fourier magnitude; the fifth, the size of the minimal voice leading to the nearest doubled subset of {0, 6}; and so on. Euclidean voice-leading distance is used for Fourier components 1–5; the "taxicab" metric is used for Fourier component 6. In all cases, voice-leading distances are calculated in continuous, unquantized pitch-class space, as described in Section I and footnote 19.

Works Cited

Callender, Clifton. 2004. "Continuous Transformations." *Music Theory Online* 10/3.
———. 2007. "Continuous Harmonic Spaces." *Journal of Music Theory* 51: 277–332.
Callender, Clifton, Quinn, Ian, and Tymoczko, Dmitri. 2008. "Generalized Voice-Leading Spaces." *Science* 320: 346–48.
Cohn, Richard. 1991. "Properties and Generability of Transpositionally Invariant Sets." *Journal of Music Theory* 35: 1–32.
Hall, Rachel and Tymoczko, Dmitri. 2007. "Poverty and Polyphony: A Connection between Music and Economics." In *Bridges Donostia: Mathematical Connections in Art, Music, and Science*, ed. Reza Sarhanghi and Javier Barallo, 259–68. Winfield, KS: Southwestern College.
Hoffman, Justin. 2008. "On Pitch-Class Set Cartography: Relations between Voice-Leading Spaces and Fourier Spaces." *Journal of Music Theory* 52:219–49.
Lewin, David. 1959. "Re: Intervallic Relations between Two Collections of Notes." *Journal of Music Theory* 3: 298–301.

———. 2001. "Special Cases of the Interval Function between Pitch-Class Sets X and Y." *Journal of Music Theory* 45: 1–29.

Quinn, Ian. 2001. "Listening to Similarity Relations." *Perspectives of New Music* 39/2: 108–58.

———. 2006. "General Equal-Tempered Harmony (Introduction and Part I)." *Perspectives of New Music* 44/2: 114–58.

———. 2007. "General Equal-Tempered Harmony (Parts II and III)." *Perspectives of New Music* 45/1: 4–63.

Robinson, Thomas. 2006. "The End of Similarity? Semitonal Offset as Similarity Measure." Paper presented at the annual meeting of the Music Theory Society of New York State, Saratoga Springs, NY.

Roeder, John. 1984. "A Theory of Voice Leading for Atonal Music." Ph.D. diss., Yale University.

———. 1987. "A Geometric Representation of Pitch-Class Series." *Perspectives of New Music* 25/1–2: 362–409.

Straus, Joseph. 2007. "Voice Leading in Set-Class Space." *Journal of Music Theory* 49: 45–108.

Tymoczko, Dmitri. 2006. "The Geometry of Musical Chords." *Science* 313: 72–74.

———. 2008. "Scale Theory, Serial Theory, and Voice Leading." *Music Analysis* 27/1: 1–49.

Vuza, Dan Tudor. 1993. "Supplementary Sets and Regular Complementary Unending Canons (Part Four)." *Perspectives of New Music* 31/1: 270–305.

Dmitri Tymoczko is a composer and music theorist who teaches at Princeton University. His book *A Geometry of Music* will be published in 2010 by Oxford University Press. He is also working on an album of pieces that combine jazz, rock, and classical styles.   〰