An Algorithm for Harmonic Analysis
Author(s): David Temperley
Source: *Music Perception: An Interdisciplinary Journal*, Vol. 15, No. 1 (Fall, 1997), pp. 31-68
Published by: University of California Press
Stable URL: http://www.jstor.org/stable/40285738
Accessed: 07/01/2015 08:48

Music Perception
Fall 1997, Vol. 15, No. 1, 31–68

An Algorithm for Harmonic Analysis

DAVID TEMPERLEY
Columbia University

An algorithm is proposed for performing harmonic analysis of tonal music. The algorithm begins with a representation of a piece as pitches and durations; it generates a representation in which the piece is divided into segments labeled with roots. This is a project of psychological interest, because much evidence exists that harmonic analysis is performed by trained and untrained listeners during listening; however, the perspective of the current project is computational rather than psychological, simply examining what has to be done computationally to produce "correct" analyses for pieces. One of the major innovations of the project is that pitches and chords are both represented on a spatial representation known as the "line of fifths"; this is similar to the circle of fifths except that distinctions are made between different spellings of the same pitch class. The algorithm uses preference rules to evaluate different possible interpretations, selecting the interpretation that most satisfies the preference rules. The algorithm has been computationally implemented; examples of the program's output are given and discussed.

In recent years, a great deal of work in music perception has focused on harmony. A number of researchers have investigated listeners' perceptions of stability and similarity relations between chords and keys, and spatial representations have been proposed to model these intuitions (Krumhansl, 1990; Lerdahl, 1988; Shepard, 1982). Others have explored the interaction of harmonic structure with other aspects of musical cognition, such as memory, expectation, and segmentation.[1] Still others have studied the role of psychoacoustics in harmony, the possibility of implementing harmonic perception using connectionist networks, the development of harmonic perception in children, and its localization in the brain.[2] Despite the important contributions of this work, however, one aspect of

1. This work is discussed further later.

2. On psychoacoustics, see Terhardt (1974) and Parncutt (1989). For a connectionist approach, see Bharucha (1987b). On development in children, see Cuddy and Badertscher (1987) and Kastner and Crowder (1990). For a review of work on brain localization and other neurophysiological work, see Zatorre (1984).

Address correspondence to David Temperley, Department of Music, Columbia University, 116th & Broadway, New York, NY 10027. (e-mail: dt3@columbia.edu)

this area has received little attention. It is generally assumed that harmony is psychologically real: at one level, a listener's processing of a piece involves dividing it into segments and labeling them as chords. But how is this done? What is the process whereby a harmonic representation is derived from a pitch representation? In this paper, I introduce a computational algorithm that I have developed for performing harmonic analysis, which may shed light on the psychological processes involved.[3]

Several assumptions of this paper should be clarified at the outset. The first concerns the psychological status of harmonic structure. It is my assumption that harmonic analysis is psychologically real for a broad population of listeners, both trained and untrained, who have exposure to tonal music. This claim is rarely made explicitly, and some might find it doubtful, but there is in fact a wealth of evidence for it. In the first place, there is ample experimental evidence—such as the studies just cited—that harmonic analysis is part of the listening process even for listeners without formal training (although of course it is often performed unconsciously). Consider, for example, Krumhansl's experiments showing that chords are perceived with varying degrees of relatedness or stability depending on the tonal context (Krumhansl, 1990, pp. 188–212); clearly, such judgments depend minimally on the chords actually being identified in some way. Other experiments establish this in a more indirect way by showing that harmony influences other aspects of perceived musical structure. For example, harmonic structure has been shown to influence segmentation, in that melodic gestures that imply V-I harmonic motion are heard as being segment endings (Palmer & Krumhansl, 1987; Tan, Aiello, & Bever, 1985). It also plays a role in expectation, in that chords that form common progressions with previous chords are expected (Schmuckler, 1989), and memory, in that melodies are more easily remembered when they can be coded in terms of alphabets built on tonal chords (Deutsch, 1982). Here again, the fact that the sequence of harmonies in a passage influences listeners' responses to it seems to indicate that the harmonies are being identified. Anecdotal evidence exists, also, for the psychological reality of harmony. Anyone who has taught music to untrained listeners knows that such listeners can often respond to cues in the music that depend on harmonic structure: for example, distinguishing major from minor and recognizing cadences (admittedly, not all students can perform these tasks all the time, but they are tasks at which many listeners have some competence). We should note, also, that the psychological reality of harmonic analysis is

3. Credit is due to a number of people for their help and advice on this project. I would especially like to thank Jonathan Kramer, Joseph DuBiel, and in particular, my dissertation advisor, Fred Lerdahl. I am also greatly indebted to Daniel Sleator, who conceived and wrote the computer implementation. For a more detailed description of the algorithm, and further discussion of many of the issues presented here, see Temperley (1996).

taken for granted by theories of higher-level musical perception and cognition that assume harmonic structure as part of the input, such as Lerdahl and Jackendoff's (1983) theory of hierarchical structures, Narmour's (1990, 1992) theory of melody,[4] and Gjerdingen's (1987) studies of musical schemata. Of course, to say that an untrained listener unconsciously performs something like harmonic analysis does not mean that his unconscious analysis of a piece is necessarily identical to that of a trained expert (and there is undoubtedly some disagreement even among experts). Still, on balance, the experimental results indicate that the analyses formed by untrained listeners are roughly similar to the "correct" analyses that experts would produce. In short, it seems reasonable for us (as trained experts) to take our harmonic analyses of pieces as indicative of the analyses that would be produced by listeners in general.

It is important to stress that no claim is being made that harmonic perception is innate; rather, the evidence suggests that it is largely learned.[5] However, it appears to be learned mainly through exposure to tonal music, rather than through explicit formal training. The fact that people must practice to learn to do harmonic analysis explicitly is no argument against the claim that they are doing it unconsciously all along (an analogy could be drawn here with syntactic or phonological analysis in language). In view of the pervasive presence of tonal harmony in Western music—not only classical music, but also hymns, carols, folk songs, show tunes, music in film, television and advertising, and so on—it should not surprise us if, as experiments seem to suggest, most listeners in Western society have acquired a substantial degree of familiarity with it.

My aim, then, has been to produce an algorithm—a rule-governed, deterministic procedure—that accurately models the process of harmonic analysis: that is, one that produces the correct harmonic analysis for a given passage of music. One might object, quite rightly, that merely finding an algorithm that correctly predicts the judgments of humans in a particular domain does not prove that humans perform the process in the same way. But it is now widely accepted in cognitive science—the work of David Marr (1982) in vision being perhaps the most notable example—that a useful way of gaining insight into psychological processes is to approach them from a purely computational point of view, asking, simply, what has to be done computationally to achieve the desired result.[6] The current project

4. Although Narmour's theory is primarily a theory of melody, harmonic factors play an important role; see, for example, Narmour (1990, pp. 212–217).

5. For a discussion of the evidence on this point, see Kastner and Crowder (1990, pp. 191–192).

6. See Marr (1982, pp. 8–38), for a discussion of his approach, especially pp. 27–29. For more general discussions of the artificial-intelligence approach to psychological problems, see Pylyshyn (1989) and Dennett (1978, pp. 109–126).

applies this same philosophy to music perception; although finding a computational model of a human process certainly does not prove that humans do it that way, the model can at least serve as a serious hypothesis for how the process might be performed, which can then be further tested in other ways, for example, through psychological experiment.

Earlier Attempts to Model Harmonic Analysis

For many musicians and certainly most theorists, performing harmonic analysis is a trivial task, requiring little thought or effort. This might lead one to suppose that the principles behind it are simple and straightforward. However, as work in other areas of psychology (e.g., speech perception and vision) has shown, tasks that are performed effortlessly by humans often prove to be highly subtle and complex. A review of some of the other studies that have addressed this issue will reveal some of the problems that arise.

The problem of harmonic analysis, as I conceive of it here, is essentially one of dividing a piece into segments and labeling each one with a root. In this sense, it is similar to traditional harmonic analysis, or "Roman numeral analysis," as it is taught in basic music theory courses. There is an essential difference here, however. In Roman numeral analysis, the segments of a piece are labeled not with roots, but rather with symbols indicating the relationship of each root to the current key: a chord marked "I" is the tonic chord of the current key, and so on. In order to form a Roman numeral analysis, then, one needs not only root information but key information. (Even once the root of a chord and the current key are known, this is not quite the same as a Roman numeral analysis, because each chord must be labeled relative to the key. However, this information is essentially determined by the root and key of each chord: if one knows that a chord is C major, and that the current key is C, the relative root of the chord can only be I.) Thus Roman numeral analysis can be broken down into two problems: root finding and key finding. My main concern here will be with the root-finding problem. In fact, however, one of the attractions of the harmonic algorithm I will propose is that it provides a basis for quite natural and powerful judgments of key; I will return to this issue later. A question arises here regarding the interaction between the root-finding and key-finding processes. It is natural to assume that key judgments are affected by root information. It is less clear whether the root-finding process can be done independently of the key-finding process, or whether some feedback is needed from key finding to root finding. I will argue that root finding can be performed effectively without using key information; the approaches I discuss here all basically share this assumption.

Several attempts have been made to devise computer algorithms that perform harmonic analysis; particularly notable are the efforts of Winograd (1968) and Maxwell (1992). Both of these algorithms begin with pitch information and derive a complete Roman numeral analysis; both root and key information must therefore be determined. I will confine my attention here to the root-finding component of the programs. Examples of the outputs of the two programs are shown in Figures 1 and 2. Both systems essentially analyze the input as a series of vertical sonorities (where any change in pitch constitutes a new sonority); the root of each sonority is determined by looking it up in a table. Simple rules are provided for guessing the identity of two-note chords (Maxwell, 1992, p. 340; Winograd, 1968, p. 20). There are then heuristics for deciding whether a sonority is a real chord or whether it is an ornamental event, subordinate to another chord (I will return to these later). This approach seems to operate quite well in the examples given; however, in many cases, it would not. Very often the notes of a chord are stated in sequence rather than simultaneously, as in an arpeggiation; neither algorithm appears capable of handling this situation. In many other cases, the notes of the chord are not fully stated at all (either simultaneously or in sequence). For example, the pitches D-F may be part of a D-minor triad, but might also be B♭ major or even G_7; as I shall show, context must be taken into account in interpreting these. (This causes problems in Winograd's example: the first chord in m. 14 is analyzed as having root D, where it should clearly be part of an arpeggiated B♭ 6/4 chord.) Problems arise also with events that are not part of any chord, so-called "ornamental dissonances" such as passing tones and neighbor notes. Both Winograd's and Maxwell's algorithms have rules for interpreting certain

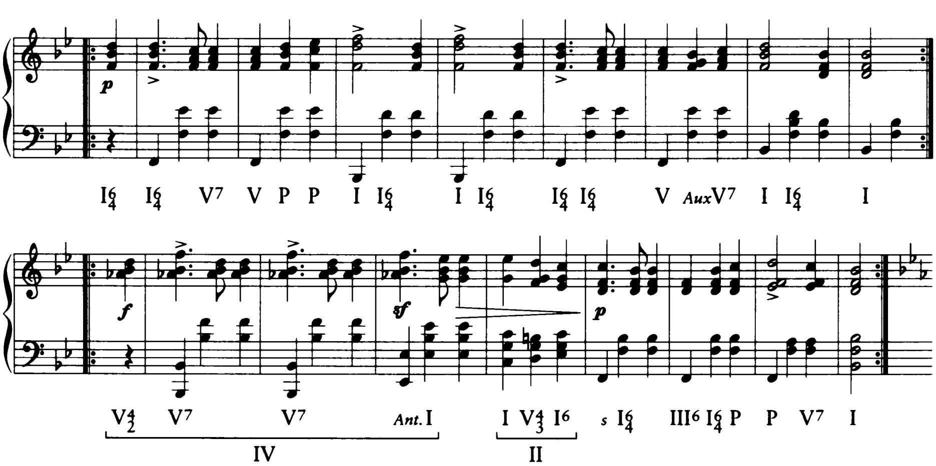

Fig. 1. Schubert, Deutsche Tänze, op. 33, no. 7. The analysis shown is the output of Winograd's harmonic analysis program. From Winograd (1968, p. 40). ©Yale University. Used by permission.

Fig. 2. Bach, French Suite no. 2 in C minor, Minuet. The analysis shown is the output of Maxwell's harmonic analysis program. From Maxwell (1992, p. 350).

verticals as ornamental, but these are not sufficient. For example, Maxwell says that any single note should be considered ornamental to the previous chord (Maxwell, 1992, p. 340). Figure 3 gives a simple example where this will not work; the A is surely not ornamental to the previous chord here.

In general, both algorithms tend to err on the side of labeling events as chordal rather than ornamental; for example, Maxwell's algorithm treats the fifth eighth note of measure 9 and the fourth eighth note of measure 11 in Figure 2 as chords, when they would usually be regarded as ornamental. A final criticism is that both programs make use of key signature and "spelling" information as part of the input; but this information would not nor-

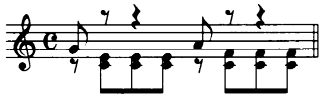

Fig. 3.

mally be available to the listener. (Maxwell's rules also rely on rhythmic notation; for example, there is a preference to have one chord change for each quarter-note beat [Maxwell, 1992, pp. 337–340].) In short, although Winograd's and Maxwell's studies contain many interesting ideas, both authors fail to address several basic problems in harmonic analysis.

Others have attempted to model harmonic perception using a neural-network or "connectionist" approach, notably Bharucha (1987b, 1991).[7] Bharucha proposes a three-level model with nodes representing pitches, chords, and keys. Pitch nodes are activated by sounding pitches; pitch nodes stimulate chord nodes, which in turn stimulate key nodes (Figure 4). For example, the C-major chord node is stimulated by the pitch nodes of the pitches it contains: C, E, and G. Bharucha's model nicely captures the intuition that chords are inferred from pitches and keys are in turn inferred from chords. The connectionist approach also offers insight into how harmonic knowledge might be acquired, an important issue that my own model does not address (Bharucha, 1987b, pp. 26–27; 1991, pp. 94–95). However, the model also has a number of problems. It was noted earlier that the approach of simply analyzing each vertical sonority one by one is insufficient. Bharucha proposes an interesting solution to this problem: a chord node is not merely activated while its pitch nodes are activated; rather, its activation level decays gradually after stimulation (Bharucha, 1987b, pp. 17–18). This might seem to offer a way of handling some of the problems encountered by the algorithms discussed earlier, such as the problem of arpeggiations; however, this solution raises other difficulties. In listening to a piece, our experience is not of harmonies decaying gradually; rather, one harmony ends and is immediately replaced by another. A similar objection could be raised to another aspect of Bharucha's model: its handling of priming or expectation. Experiments have shown that, when listeners hear a chord, they are primed to hear closely related chords (e.g., they respond more quickly to related chords than to unrelated ones in making judgments of intonation). Bharucha's model attempts to handle this by allowing the key nodes stimulated by chord nodes to feed back and activate the nodes of related chords (Bharucha, 1987b, pp. 18–21). The problem here

7. Two other less ambitious attempts to model tonal harmony using neural networks are Scarborough, Miller, and Jones (1991) and Laden and Keefe (1991).

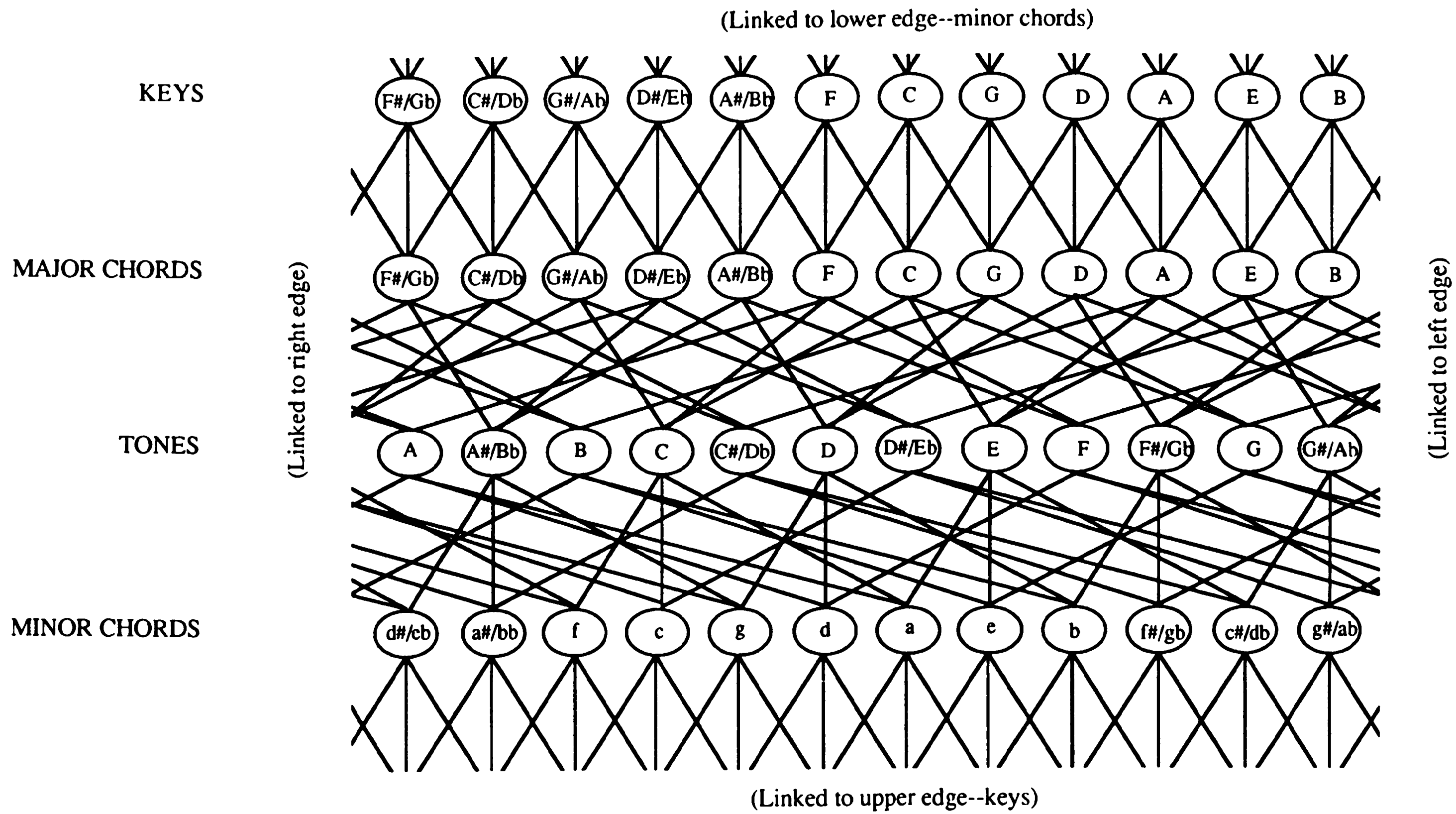

Fig. 4. Bharucha's connectionist model, showing relationships between pitches, chords, and keys. The ovals represent nodes; the lines represent connections between them. (From Bharucha, 1987a. Copyright by the Cognitive Science Society.)

is this: what exactly does the activation of a chord node represent? One would assume that it represents the chord that is actually being perceived at a given moment. But now Bharucha is suggesting that it represents something quite different, the amount that a chord is primed or expected. In fact, the idea of priming is a very important one, but the degree to which a chord is heard is different from the degree to which it is expected. I will offer an alternative approach to these problems.

A very different approach to harmonic perception is taken by Parncutt (1989). The three models discussed so far all assume that harmonic perception begins with pitch. Recovering chord and key information from music requires forming an accurate representation of the pitches; this then serves as the basis for further processing. Parncutt's work challenges this assumption. Parncutt argues that many aspects of musical cognition depend not on pitches as they occur in the score, but rather on "virtual pitches."[8] A musical pitch is made up of a combination of sine tones or pure tones: a fundamental plus many overtones. But the overtones of a pitch may also be understood as overtones of different fundamentals. For example, if one plays C_4-E_4-G_4 on the piano, some of the components of E_4 and G_4 are also overtones of C_4; others are overtones of pitches that were not even played, such as C_3 and C_2. In this way, a set of played pitches may give rise to a set of "virtual pitches" that are quite different in frequency and strength. Parncutt uses virtual pitch theory to make predictions about a number of aspects of musical cognition, such as consonance levels of chords and the number of perceived pitches in a chord; it is also used to predict the roots of chords. The root of a chord, Parncutt proposes, is the virtual pitch that is most strongly reinforced by the pure-tone components of the chord (Parncutt, 1989, pp. 59, 139–142). The theory's predictions here are quite good for complete chords (such as major and minor triads and sevenths). They are less good for incomplete chords; for example, the root of the dyad C-E♭ is predicted to be E♭ (Parncutt, 1989, pp. 146–150), as opposed to C or A♭. In cases in which consonance levels or roots of chords are not well explained by his theory, Parncutt suggests that they may have "cultural rather than sensory origins" (Parncutt, 1989, p. 142).

The psychoacoustic approach to harmony yields many interesting insights. However, it is rather unsatisfactory that, in cases where the theory does not make the right predictions, Parncutt points to the influence of cultural conditioning. This would appear to make the theory unfalsifiable; moreover, it is certainly incomplete as a theory of root judgments, because some other component will be needed to handle the "cultural" part. But even if Parncutt's theory were completely correct as far as it went, in a certain sense it goes no further than the other studies explored here in ac-

8. The idea of virtual pitches was first formulated by Terhardt (1974).

counting for harmonic perception. It accounts for the fact that certain pitch combinations are judged to have certain roots, and it offers a more principled (although imperfect) explanation for these judgments than other studies we have seen. But as we have noted, root analysis involves much more than simply going through a piece and choosing roots for a series of isolated sonorities. One must also cope with arpeggiations, implied harmonies, ornamental dissonances, and so on. A psychoacoustic approach does not appear to offer any solution to these problems. This is not to say that psychoacoustics is irrelevant to harmony; clearly it is not (indeed, it might be incorporated into my own approach in a limited way, as I will discuss). But it seems highly problematic to try to explain harmonic perception solely in terms of psychoacoustic principles.

Although these studies contain many valuable ideas, none of the studies offers a satisfactory solution to the problem of harmonic analysis. Some of these models also suffer from being highly complex. Maxwell's program (the chord-labeling component alone) has 36 rules; Winograd's program, similarly, has a vast amount of information built into it (as can be seen from his article). Bharucha's and Parncutt's models are more elegant; however, they seem even less adequate than Maxwell's and Winograd's systems in handling the subtleties of harmonic analysis—ornamental dissonances, implied harmonies, and the like. I now propose a rather different approach to these problems.

Spatial Representations and the "Line of Fifths"

Like the other studies discussed earlier (with the exception of Parncutt's), the algorithm I propose begins with a representation showing pitch information. Essentially, the input I assume is a two-dimensional representation, with pitch on one axis and time on the other, similar to a "piano roll" (an example is shown in Figure 5).[9] Pitch events in the input representation are categorized into chromatic scale steps, reflecting the well-established fact that pitch perception is "categorical" in nature. Pitch events are also labeled by pitch class (this would seem to be a simple matter, once their chromatic scale step is known), but no further information is provided about them. In particular, the input representation does not show the correct "spelling" of each note, for example, A♭ versus G♯ (in contrast to Winograd's and Maxwell's programs, which were given this information);

9. In beginning with such a representation, I do not wish to suggest for a moment that the process of deriving pitch information from sound input is a minor or trivial stage in perception; clearly it is not, and it has itself been the subject of considerable study (see Tanguiane [1994] for a review). However, this is not our concern here.

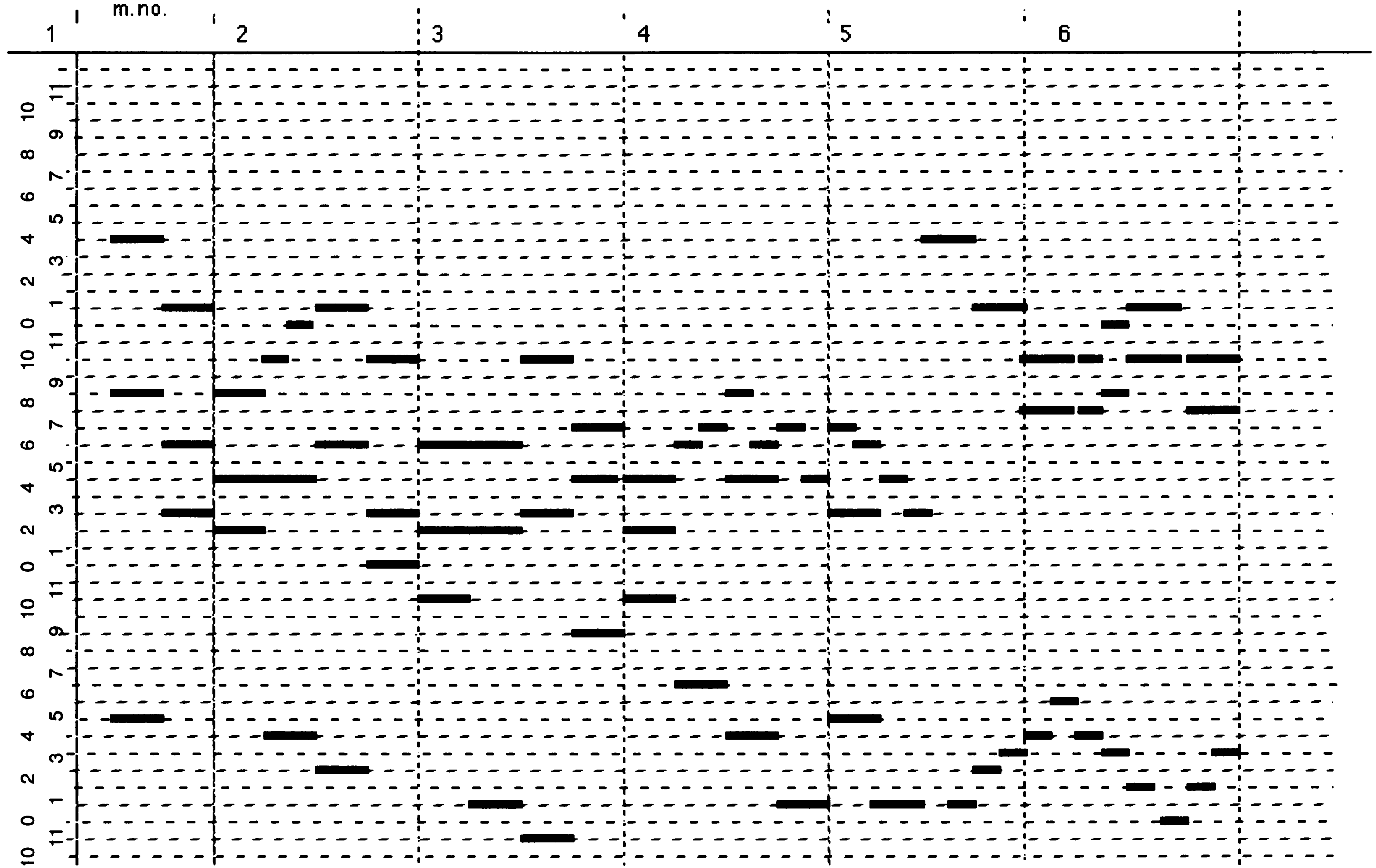

Fig. 5. A pitch-time representation of the opening of the Gavotte from Bach's French Suite no. 5.

this must be determined by the algorithm, as I will discuss. Given such an input, the algorithm must form a harmonic representation in which the piece is completely divided into segments or "chord spans," each one labeled with a root. Before we turn to the algorithm itself, we must consider one further issue: the kind of spatial representation that will be used.

A good deal of work on pitch and harmony has involved spatial representations. This includes work by music theorists, from Riemann and Schönberg to the more recent work of Lerdahl (1988, 1992, 1996); it also includes work by psychologists, notably Shepard (1982) and Krumhansl (1990), whose models have been based largely on experimental results. Most of this work is not directly relevant to our purposes here, as it involves intuitions and experimental judgments about relationships between harmonic entities once they are formed, rather than the process of identifying them. I will argue, however, that spatial representations are of great importance in chord labeling. Two examples will illustrate this point. Consider the two short passages in Figure 6. The final chord of each passage, C-E, might either be C major or A minor. Probably it would be interpreted as C major in the first case, A minor in the second case; why is this? One possibility is that chords are mentally represented in some spatial way, and we prefer to label chords as being close to previous chords on the space. Various models might be used for this purpose, but the one I propose is an extremely simple one: a "line of fifths," in which roots are arranged by fifths, similar to the circle of fifths, but extending infinitely in either direction (Figure 7). It can be seen that such a model might allow us to resolve the ambiguity in Figure 6. In a passage where the first two chords are C and G, a third root of C will be closer to the previous roots than a root of A. However, if the first two roots are A and E, a third root of A will be closer. Thus C will be preferred in the first case, A in the second.

Now consider Figure 8, which shows a slightly different situation. Here, the root of the first measure is clearly G, followed by C; what is the most likely interpretation of the final chord? In the first passage, it seems to me

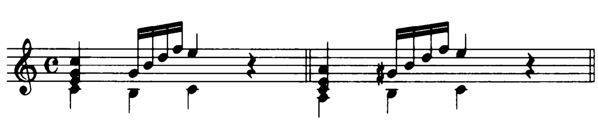

Fig. 6.

...Bbb Fb Cb Gb Db Ab Eb Bb F C G D A E B F# C# G# D# A# E# B# Fx Cx...

Fig. 7. The "line of fifths."

Fig. 8.

that the probable interpretation is as A_7; in the second, it is as E♭$_7$. Notice that, in this case, we cannot use chord distances to resolve the ambiguity, because the first two chords are the same in both cases. I propose an alternative solution. Let us suppose that not only roots, but pitches as well, are represented on the line of fifths, so that we distinguish between, for example, C♯ and D♭. This adds a further stage to the process: before harmonic analysis can begin, each pitch must be mapped on to a position on the line of fifths. The line-of-fifths position of an event is constrained by its pitch class—an event of pitch class 0 can be either C or B♯, but not F♯ or G—but other considerations must be taken into account to decide between the possibilities. (I call these different spellings of the same pitch class "tonal pitch classes," as opposed to the conventional 12-category system of "neutral pitch classes.") The main consideration here is the same as that stated earlier with regard to roots: attempt to label events so that they are close together on the line of fifths. Returning to Figure 8, how would the C♯/D♭ in each passage be labeled? In the first passage, C♯ is closer to previous events on the line of fifths; in the second, however, D♭ is closer. (Getting this result will depend on exactly how the "closeness" of events is calculated.) Using this simple rule, then, we could arrive at the correct spelling for the pitch events in this passage. But now, we further assume that the harmonic representation takes the spelling of pitches into account, so that the chord G-C♯ can have a root of A but not E♭, but the chord G-D♭ may have a root of E♭ but not A. By this process, then, we can arrive at the desired harmonic interpretation for each passage.

The line-of-fifths model brings up an important, and much neglected, issue in music perception. Most models of pitch perception have been based on neutral pitch classes: for example, Shepard's (1982) various music spaces, Krumhansl's "key profile" (Krumhansl, 1990, pp. 21–31), and Bharucha's connectionist model (discussed earlier) all label pitches in terms of neutral pitch class, without any further subcategorization.[10] However, this is not

10. One important exception here is Longuet-Higgins (1962), who argues for the psychological importance of different pitch spellings. However, his model, which is based on psychoacoustics, also posits distinctions even between pitches of the same spelling (D a third below F is different from D a fifth above G); clearly, this is rather different from what I propose here.

the prevailing assumption in tonal music theory and notation; there, distinctions are commonly made between (for example) the pitches A♭ and G♯, the chords A♭ major and G♯ major, and so on. Which model is correct from a psychological point of view, the neutral-pitch-class (NPC) model or the tonal-pitch-class (TPC) model? It is my view that the TPC system is strongly preferable. One could argue, first of all, that TPC distinctions are experientially real and important in and of themselves. An E♭ and a D♯ next to each other simply sound like different pitches (even on a piano); A♭ major seems closer to C major than G♯ major does (both in terms of chords and keys). However, using a TPC model is also more convenient and effective even in terms of making correct distinctions between *neutral* pitch classes. Figure 8 offered one example; here, TPC distinctions allow us to correctly label the final chord as A in one case, E♭ in the other. There are other advantages as well, as I will discuss later.

The idea, then, is to locate both pitches and roots so that they are maximally close together on the line of fifths. How exactly is this to be accomplished? One simple possibility is to spell each event (pitch or chord) so that it is maximally close to the previous event. Further thought shows that this is not sufficient, however. Consider the pitch sequence A-B-C♯-D-G♯. The final event should be spelled as G♯ rather than A♭; but these two TPCs are equally close to the previous TPC, D (both are six steps away on the line of fifths). Rather, it seems that the current event should be labeled to maximize its closeness to *all* previous events, with more-recent events being weighted more than less-recent ones. In the current model, a "center of gravity" is taken, reflecting the average position of all prior events on the line of fifths (weighted for recency); the new event is then spelled so as to maximize its closeness to that center of gravity.[11]

A further point is needed regarding my use of the line-of-fifths model. Another alternative would be to use a multidimensional chord space such as that proposed by Krumhansl and Kessler (1982) and further developed by Lerdahl (1988), shown in Figure 9. This space shows the seven diatonic chords of a key. One axis represents the diatonic circle of fifths, the other the diatonic circle of thirds; the space wraps around in both dimensions.

11. The elegance of this solution is another point in favor of the "line-of-fifths" model. One might also use a wraparound space such as the circle of fifths for resolving root ambiguities. (There would be no point in using it for pitch labeling, because spelling differences are not represented.) However, it is quite unclear how such a "center of gravity" is to be calculated. Numbering the points in the space and taking the average will not work; the results will depend entirely on how the points are numbered. If C is 0 and A♭ is 4, the center of gravity between them is 2, which is B♭; however, if A♭ is 0 and C is 8, the center of gravity is 4, which is E. The same applies to multidimensional wraparound spaces, such as Lerdahl's (discussed later).

vii°	ii	IV	vi	I	iii	V
iii	V	vii°	ii	IV	vi	I
vi	I	iii	V	vii°	ii	IV
ii	IV	vi	I	iii	V	vii°
V	vii°	ii	IV	vi	I	iii
I	iii	V	vii°	ii	IV	vi
IV	vi	I	iii	V	vii°	ii

Fig. 9. Lerdahl's "chordal space." From Lerdahl (1988, p. 326).

However, there is a problem with this model for our purposes. The space shown in Figure 9 is a "within-key" space; according to Lerdahl's theory, there are different chord spaces for each key. Not all chords are shown in any one chord space; moreover, any given chord will be represented in various different spaces (e.g., C major is I/C, V/G, and so on). This means that one must know the key one is in in order to calculate the distance between any two chords; in effect, it assumes strong "top-down" influence from the key-finding level to the root-finding level. This presents a complication, one that I believe is unnecessary for the purpose of root finding. However, the possibility of using other spaces for chord finding should be explored further (see Temperley, 1996, pp. 62–74, 143, for discussion).

The basic scheme that is emerging is as follows. Before beginning the process of harmonic analysis, the algorithm chooses a TPC label for each pitch event; in so doing, it maps each event on to a point on the line of fifths. This is the TPC level of the algorithm. The algorithm then proceeds to the harmonic level, where it divides the piece into segments labeled with roots. At this stage, too, it maps roots on to the line of fifths, attempting to choose roots so that the roots of nearby segments are close together on the line. Thus the line-of-fifths model serves several purposes. It allows the spellings of pitches to be determined; in the case of roots, it not only selects spellings, but also resolves ambiguities such as those in Figures 6 and 8. The harmonic level involves other considerations as well, however, as I will explain. The basic framework of the algorithm is shown in Figure 10 (note that metrical structure is also required as input; this will be discussed). I will now give a more detailed overview of the algorithm, using an example: the Gavotte from Bach's French Suite no. 5 in G major, shown in Figure 11.

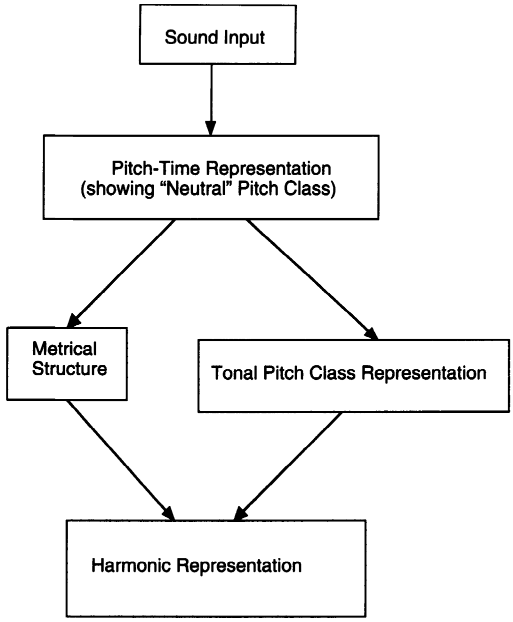

Fig. 10.

The Rules of the Algorithm

At a conceptual level, the algorithm consists of a set of "preference rules." Preference rules, which were first used in Lerdahl and Jackendoff's *Generative Theory of Tonal Music,* are rules governing the formation of some kind of structure or representation, stating the criteria for preferring some representations over others (Lerdahl & Jackendoff, 1983, pp. 9, 39–43). When there are multiple preference rules, they may interact in complex ways, sometimes supporting each other and sometimes conflicting; the preferred representation is the one that is most favored, on balance, by all the rules together.

As discussed earlier, the algorithm's input is a "pitch-time" representation, showing the pitches of a piece arranged in time. Such a representation is shown in Figure 5, for the beginning of the Bach Gavotte. The algorithm's first step is to map each of these pitches on to the line of fifths, thereby creating the "TPC representation." This can be thought of as a two-dimensional representation, with time on one axis and the line of fifths on the other, with each pitch represented as a line segment on the plane, as in Figure 12. Consider the first three chords (nine notes) of the Bach. These pitches could be spelled and as shown in the score: G-D-B-G-B-G-F#-A-D;

Fig. 11. Bach, French Suite no. 5, Gavotte.

alternatively, they could be spelled G-D-B-F♯♯-C♭-G-G♭-A-E♭♭. The first spelling is clearly preferable, but why? The main consideration here has already been stated: try to label events so that they are close together on the line of fifths. The first way of spelling the pitches locates them very close together

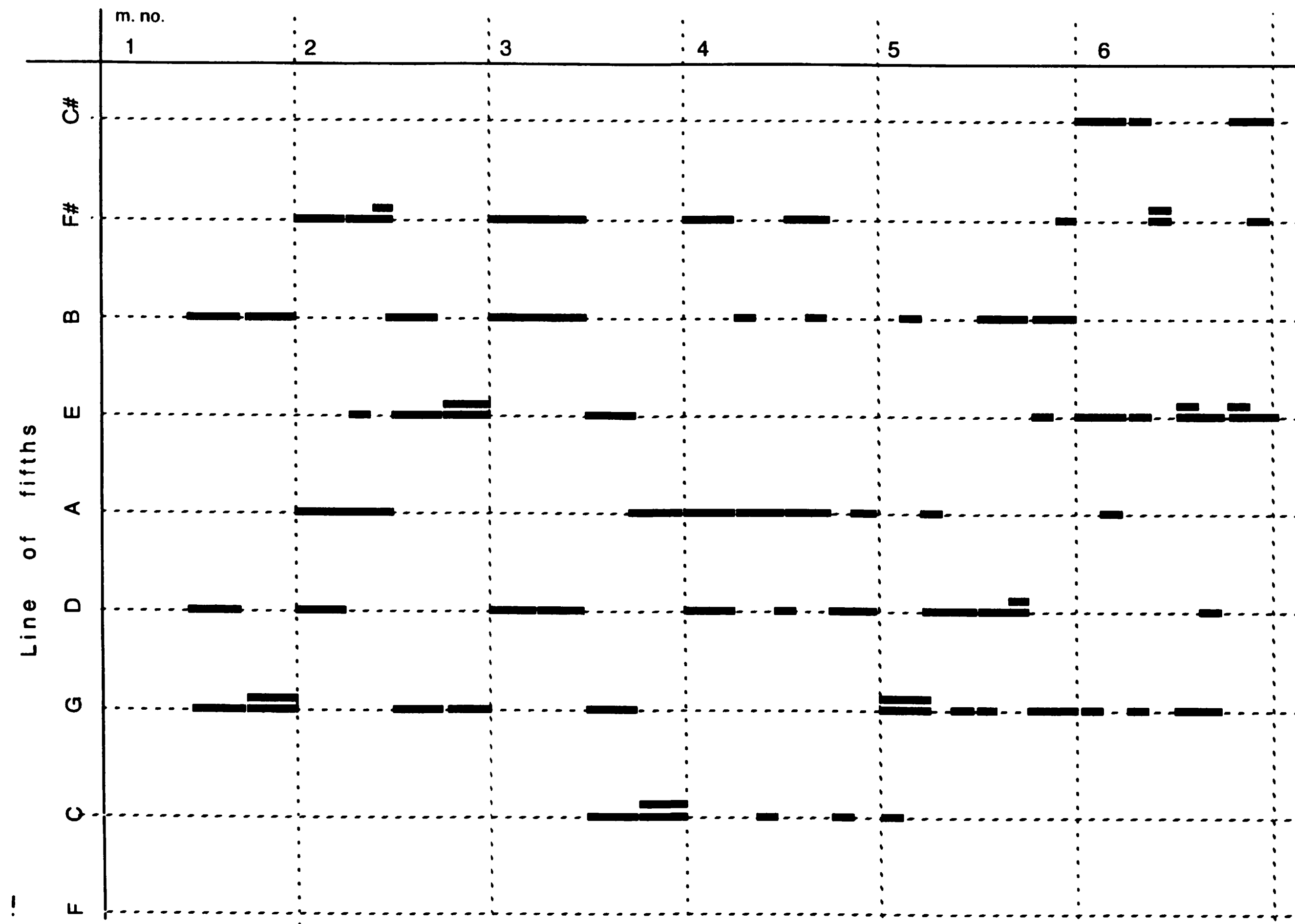

Fig. 12. The tonal-pitch-class representation of the opening of the Bach Gavotte.

on the line; the second way leaves them much less "closely packed." We state this rule as follows:

Pitch Variance Rule: Try to label nearby pitches so that they are close together on the line of fifths.

Note that the rule applies to nearby pitches (pitches that are close together in time). For pitches within a few seconds of each other, the pressure is great to locate them close together on the line; for pitches widely separated in time, the pressure is much less. One problem with the rule should be mentioned. Although it chooses correctly between the two possible spellings of the Bach chords just given, what about a third alternative: A𝄫-E𝄫-C♭-A𝄫-C♭-A𝄫-G♭-B𝄫-E𝄫? This is identical to the correct spelling, except that *all* the events are shifted over 12 steps on the line. The events here are as closely packed as they are in the correct spelling. To resolve this problem, the algorithm automatically assigns the first event in the piece to a certain cycle of the line of fifths (the region between F♯ and D♭); once this is done, the spelling of subsequent events is determined by the pitch variance rule.

As explained earlier, the formalization of the pitch variance rule depends on the idea of a "center of gravity" (COG). For each pitch event, a COG is calculated, reflecting the average position of all previous pitches on the line of fifths, with more recent pitches weighted more than less recent ones (this assumes that the spelling of all previous pitches has already been determined). Pitch events are weighted for duration here, so that longer events affect the COG more. The algorithm then attempts to spell the new pitch in such a way that it is maximally close to this center of gravity. The way in which the TPC representation is generated has several complications, however, which I explain later.

Once the TPC representation is complete, the algorithm creates the "harmonic representation." Here, the piece is divided into segments, or "chord spans," labeled with roots; each root is a point on the line of fifths. Again, we can imagine a two-dimensional representation; this time, line segments represent chord spans rather than pitches (Figure 13 shows such a representation for the opening of the Gavotte). For each chord-span, a root must be selected (the segmentation of the piece into chord spans must also be determined; I will discuss this later). One important factor is clearly the pitches that each span contains. Let us consider the second chord of measure 1: G-B-G. Simply considering the segment out of context, we know that its root is unlikely to be D or F; the most likely root is G, because both G and B are chord tones of the G major chord. E is also a possibility, but even considering this segment in isolation, G would seem more likely. The way we capture these intuitions is as follows. Every TPC has a relationship to every root, depending on the interval between them. The TPC G is 1 of

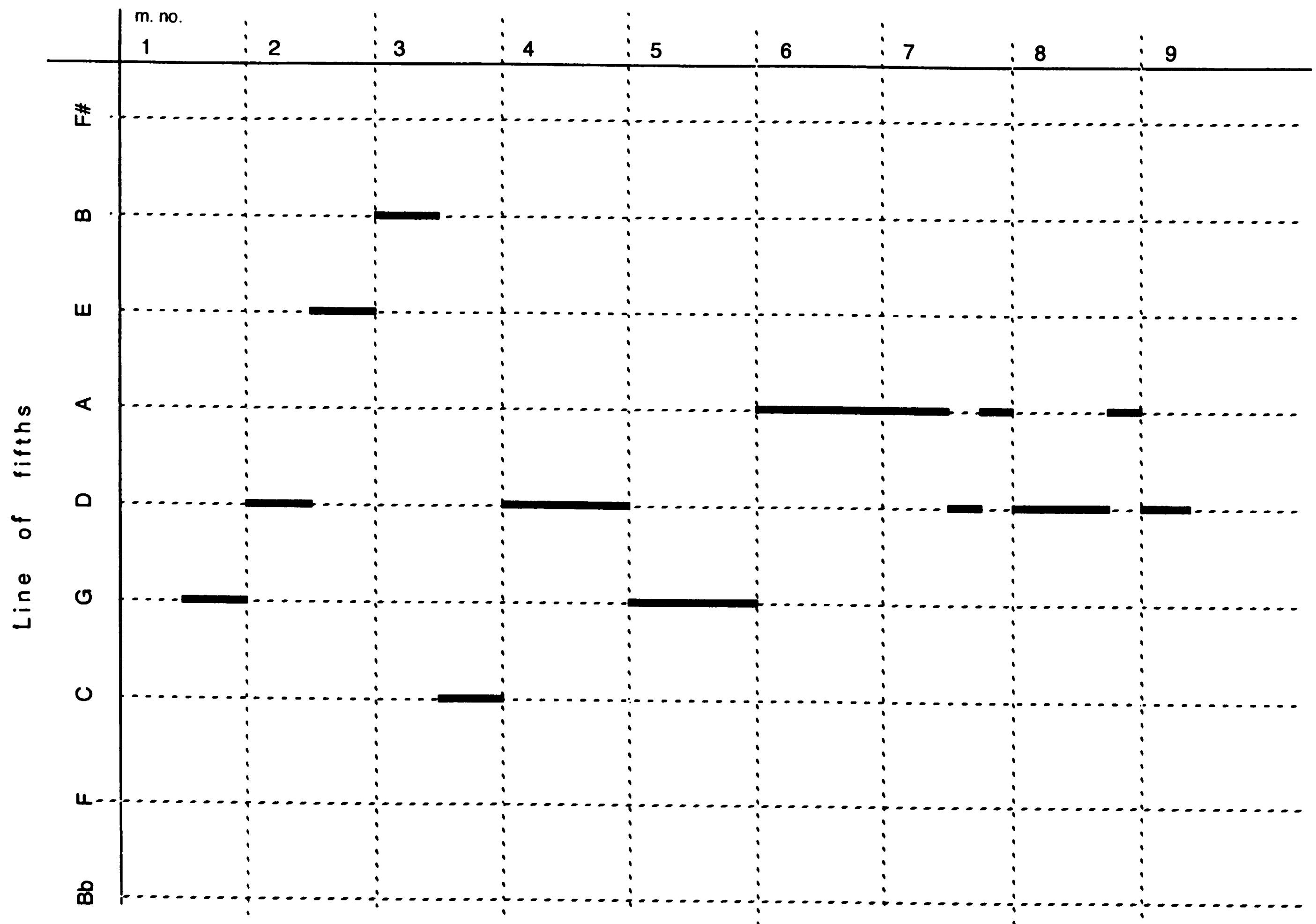

Fig. 13. The harmonic representation of the opening of the Bach Gavotte.

G, 5 of C, 3 of E♭, ♭3 of E, and ♭7 of A. (Other relationships might also be added, but we will consider only these five here.)[12] Certain relationships are more preferred than others; we try to choose roots for each segment so that the relationships created are as preferred as possible. We call this the compatibility rule, stated as follows:

Compatibility Rule: In choosing roots for chord spans, prefer certain TPC-root relationships over others. Prefer them in the following order: 1, 5, 3, ♭3, ♭7, ornamental. (An ornamental relationship is any relationship besides these five.)

Given the chord G-B-G, this rule will prefer a root choice of G over E. A root of G will result in TPC-root relationships of 1, 1, and 3, whereas a root of E will result in ♭3, ♭3, and 5; the former choice is clearly preferred by the compatibility rule. Roots that involve "ornamental" relationships—those other than the ones specified—are still less preferred. (Note that the compatibility rule considers TPCs, not NPCs. The importance of this has already been discussed; for example, it allows the algorithm to make the correct root choices in Figure 8. This has other advantages as well, as I will explain.)

Note that the algorithm only labels each chord span with a root. In this sense it differs from conventional harmonic analysis (aside from the fact, discussed earlier, that conventional analysis identifies roots in relative rather than absolute terms). Roman numeral analysis gives other information about chords as well, such as their mode (major or minor) and extension (triad, seventh, etc.). However, such information appears to be fairly easily accessible. I will return to this point.

Now consider the second quarter of measure 2. The E is clearly ornamental (how this is determined will be explained later); the chord tones of the segment are then F♯-A-F♯ (the A is held over from the previous beat). The correct root here is D, but the compatibility rule alone does not enforce this. A root of D will yield TPC-root relationships of 3-5-3, whereas a F♯ root will yield 1-♭3-1; the compatibility rule does not express a clear preference. Consider also the last quarter of measure 2; here the pitches are E-G-E, offering the same two interpretations as the previous case (3-5-3 vs. 1-♭3-1). But here, the root is E; in this case, then, the 1-♭3-1 choice is preferable. Clearly, another rule is needed here. The rule I propose is a very simple one: we prefer to give each segment the same root as previous or following segments. In this case, the first beat of measure 2 clearly has root D; there is then strong pressure to assign the second beat the same root as well.

12. The reason why these five relationships are the most preferred ones (and why some are more preferred than others) is a complex issue, which I will not explore here; psychoacoustic factors are undoubtedly relevant (see Parncutt, 1989, discussed earlier).

Another way of saying this is that we prefer to make chord spans as long as possible (where a chord span is any continuous span of music with a single root). This rule—which we could tentatively call the "long-span" rule—also addresses another question: we have been assuming segments of a quarter note, but why not consider shorter segments such as eighth-note segments? For example, what is to prevent the algorithm from treating the third eighth note of measure 5 as its own segment and assigning it root D? Here again, the "long-span" rule applies; spans of only one eighth note in length (that is, an eighth-note root segment with different roots on either side) will generally be avoided, although they may occasionally arise if no good alternative is available.

Although it is true that long spans are usually preferred over shorter ones, further consideration shows that this is not really the principle involved. Consider Figure 14: The first note of measure 2 could be part of the previous F segment (as a 1), or it could be part of the following G segment (as a ♭7). The compatibility rule would prefer the first choice, and the long-span rule stated earlier expresses no clear preference; why, then, is the second choice preferable? The reason is that we do not simply prefer to make spans as long as possible; rather, we prefer to make spans start on strong beats of the meter. This has the desired effect of preferring longer spans over shorter ones (strong beats are never very close together; thus any very short span will either start on a weak beat itself, or will result in the following span starting on a weak beat). In measure 2 of the Bach, for example, having spans start on strong beats will mitigate against starting spans on the second and fourth quarter notes, as these are relatively weak beats. However, this has the additional effect of aligning chord-span boundaries with the meter, thus taking care of cases like Figure 14. We express this as the "strong-beat rule."

Strong-Beat Rule: Prefer chord spans that start on strong beats of the meter.

The strong-beat rule raises a complication: it means that the algorithm requires metrical structure as input. The kind of metrical structure I am assuming is that proposed by Lerdahl and Jackendoff (1983): a structure

Fig. 14.

consisting of a series of levels of evenly spaced beats, with every second or third beat at one level being a beat at the next level up. (The way this structure is derived is a complex cognitive process in itself, but this is not our concern here; see Lerdahl & Jackendoff, 1983, pp. 68–103.) Every level of beats has a certain time interval associated with it, which is the time interval between beats at that level. According to Lerdahl and Jackendoff, every time point in a piece that is the beginning or ending of a note must coincide with a beat (Lerdahl & Jackendoff, 1983, p. 72), although there may also be beats that do not coincide with note beginnings or endings. Thus each such time point has a certain highest beat level, that is, the highest beat level at which that time point is a beat; and each such time point has a metric strength, which is given by the time interval of its highest beat level. A time point whose highest beat level has a long time interval is metrically strong; the longer the time interval, the stronger the beat. This, essentially, is how the rule above is expressed formally; there is a preference for starting chord spans on metrically strong beats. (Time points that are not beats of the metrical structure at all are simply disallowed as segment boundaries. This seems to conform to intuition; it also seems to be a logical extension of the strong-beat rule and greatly limits the possibilities that the algorithm must consider.)

A further rule is nicely illustrated by measure 16 of the Bach Gavotte. If we assume for the moment that the C and A in the right hand and the F♯ and A in the left hand are ornamental dissonances (again, this will be explained later), this leaves us with chord tones of G and B. The compatibility rule would prefer a root of G, but E seems a more natural choice; why? This brings us to the consideration discussed earlier: we prefer to choose roots that are close together on the line of fifths. The previous span clearly has root B; we therefore prefer E over G as the root for the following span. The same applies to the first half of measure 19; the chord tones here, C and E, could have a root of A or C, but because the previous span has root G, C is preferable (in this case, the compatibility rule reinforces this choice). We express this rule as follows:

Harmonic Variance Rule: Prefer roots that are close to the roots of nearby segments on the line of fifths.

The implementation of this rule is similar to that for the pitch variance rule. For each new span, a COG is calculated, reflecting the average position of all previous roots on the line of fifths, weighted for recency; for the new span, roots closer to this COG are preferred.

The final rule of the algorithm concerns ornamental dissonances. We have been assuming that certain events are ornamental. This means that, in the process of applying the compatibility rule (i.e., looking at the pitches in a segment and their relationship to each root), certain pitches can simply be neglected. But how does the algorithm know which pitches can be ornamental? The key to our approach here is an idea proposed by Bharucha (1984). Bharucha addressed the question of why the same pitches arranged in different orders can have different tonal implications: B-C-D♯-E-F♯-G has very different implications from G-F♯-E-D♯-C-B, the same sequence in reverse. (Bharucha verified this experimentally, by playing subjects each sequence followed by either a C-major or B-major chord. The first sequence was judged to go better with C major, the second with B major [Bharucha, 1984, pp. 497–501].) He hypothesized what he called the "anchoring principle": a pitch may be ornamental if it is closely followed by another pitch a step or half-step away.[13] In the first case, all the pitches may be ornamental except C and G; in the second case, D♯ and B may not be ornamental. It is then the nonornamental pitches that determine the tonal implications of the passage. The algorithm I propose applies this same principle to harmonic analysis. The algorithm's first step is to identify what I call "potential ornamental dissonances" (hereafter PODs). A POD is an event that is closely followed by another pitch a step or half-step away in pitch height. What is measured here is the time interval between the *onset* of each note and the onset of the next stepwise note. For example, the first E in the melody in measure 2 is a good POD because it is closely followed by F♯; the A in the melody in measure 5 is closely followed by G. However, the G in measure 5 is not closely followed by any pitch in a stepwise fashion; it is not a good POD. (The "goodness" of a POD is thus a matter of "more or less" rather than "all or nothing.") The algorithm then applies the compatibility rule, considering the relationship between each TPC and a given root. As mentioned earlier, if the relationship between an event's TPC and the chosen root is not one of the "chord-tone" relationships specified in the compatibility rule—1, 5, 3, ♭3, or ♭7—that event is then ornamental. Any pitch may be treated as ornamental, but the algorithm prefers events that are good PODs. We express this rule as follows:

Ornamental Dissonance Rule: An event is an ornamental dissonance if it does not have a chord-tone relationship to the chosen root. Prefer ornamental dissonances that are closely followed by an event a step or half-step away in pitch height.

The algorithm satisfies this rule in an indirect fashion—not by labeling notes as ornamental once the root is chosen (this follows automatically), but by choosing roots so that notes that emerge as ornamental are closely followed in stepwise fashion. However, there is always a preference for

13. Bharucha actually expresses this in terms of scales: an anchored pitch is one that is followed by another pitch a step away in the current diatonic scale (Bharucha, 1984, pp. 494–495). Because the algorithm has no representation of scales, this option is not available to us here.

considering events as chord tones rather than ornamental dissonances, even if they are good PODs (this is specified in the compatibility rule).

The "anchoring principle" does a good job of identifying a variety of kinds of ornamental dissonances. It handles ordinary passing tones (such as the eighth-note E in measure 2), and neighbor notes (the D in the left hand in measure 6), as well as unprepared neighbors and appoggiaturas, notes that are followed but not preceded by stepwise motion (such as the C in measure 5). It also handles "double neighbors," such as the C-A in measure 16: a pair of ornamental tones on either side of a following chord tone. The C is considered a (fairly) good ornamental dissonance because it is followed (fairly) closely by the B; the fact that an A is in between is irrelevant. However, not all kinds of ornamental dissonances are captured by this rule. One important exception is escape tones, such as the F♯ at the end of measure 8; another example is anticipations, ornamental tones (often at the end of a measure) that are followed by another note of the same pitch, such as the G at the end of measure 24. The current version of the algorithm cannot handle such notes, although in principle it should be possible to incorporate them.

The Basic Operation of the Algorithm

The four rules just presented comprise the entire harmonic level of the algorithm. We will now examine some other issues in the way the algorithm operates and the way the rules are applied. We will assume for now that the TPC representation has already been completed (although the situation is in fact more complex, as I will discuss); the algorithm's task, then, is to generate the harmonic representation.

Let us begin with an idealized (but unworkable) procedure for the algorithm. In this procedure, the algorithm generates all possible "well-formed" harmonic interpretations for an entire piece, where a well-formed interpretation is simply one in which the piece is exhaustively segmented into chord spans, each one labeled with a single root. Let us suppose that the piece is divided into very short segments, each of which must be assigned a root; a "chord span" then simply emerges as a series of contiguous segments with the same root. The algorithm gives each interpretation a numerical score and then chooses the one with the highest score. Each preference rule assigns a score to each interpretation; the total score for the interpretation is the sum of the scores for the four rules. The way the scores are calculated will not be discussed in detail here; I will simply present the basic ideas. For the compatibility rule, each TPC-root relationship has a certain score (with more preferred relationships having higher scores); thus each pitch event in an interpretation yields a certain score, depending on the relationship of its

TPC to the current root. The compatibility score for an interpretation is the sum of all the compatibility scores for the notes in the piece, given the roots of the interpretation. (As with the pitch variance rule, notes are weighted for duration here, so that longer events affect the results more.) For the harmonic variance rule, each segment is assigned a penalty, which is a function of the distance of the segment's root from the current COG on the line of fifths; segments that are farther from the COG receive higher penalties. For the strong-beat rule, a penalty is applied to segments whose root differs from the root of the previous segment (so that only the first segment of each chord span receives a penalty). The penalty is determined by the strength of the beat on which it begins; segments beginning on stronger beats receive lower penalties. For both the variance rule and the strong beat rule, the total score for the interpretation is the sum of the scores for all the segments in the interpretation. For the ornamental dissonance rule, each event that is ornamental (i.e., that is not a chord tone relative to the current root) receives a penalty depending on how "good" it is as an ornamental dissonance, with better ornamental dissonances receiving lower penalties; the total score for the interpretation sums all these scores. Thus every complete interpretation of a piece receives a single global score that indicates how good an interpretation it is, based on (a) how compatible the root of each span is with the pitches the span contains, (b) how close together the roots are on the line of fifths, (c) how well the span boundaries are aligned with the metrical structure, and (d) the "goodness" of any ornamental dissonances that are entailed by the interpretation. The numerical values for these rules are of course very important; they indicate not only how the penalties for a given rule vary under different conditions (e.g., how preferable a TPC-root relationship of 1 is relative to 5; how preferable a 1-sec ornamental dissonance is compared with a 2-sec one), but also how much weight each rule carries relative to the others.

It should be clear that this procedure for the algorithm is neither computationally practical nor psychologically plausible. Each segment has an infinite number of possible roots (the root could be at any position on the line of fifths), and the possible ways of combining these segment interpretations will go up exponentially with the number of segments. One might ask, why must the algorithm consider complete interpretations of the piece; why not analyze the piece one segment at a time? The reason is that, as we have seen, the correct analysis of a segment depends on its context. Because of the variance rule, the preferred root of a segment may depend on preceding segments.[14] The first half of measure 16, considered in isolation, might

14. Context is also important because of the strong-beat rule: the penalty assigned to a segment depends on whether the previous segment has the same root. However, this is less difficult to handle computationally.

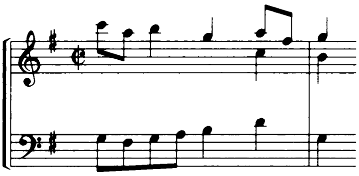

Fig. 15. A revised version of measure 16 of the Bach Gavotte.

well be analyzed as having root G; it is only the context that makes us prefer root E (specifically the fact that E is closer to the root of the previous span). For that matter, the interpretation of a segment might also be affected by later segments. Suppose the first half of measure 16 was followed instead by an unambiguous G-major chord (as shown in Figure 15); this might well cause us to hear the first half of measure 16 as having root G as well. This is accounted for nicely by the current approach; hearing the first half of measure 16 as having root G would give the following segment a better score on both the variance rule and the strong-beat rule; overall, then, this interpretation might well be preferred. But this further points up the importance of looking at total interpretations rather than isolated-segment interpretations. The preferred interpretation of a segment of the piece, therefore, is not the segment that receives the highest score in isolation, but rather *the segment interpretation that is part of the highest scoring overall interpretation.* For this reason, the basic approach that I have described—choosing the best total interpretation among all possible ones—is, in principle, necessary. But the procedure I have described for it, generating all interpretations and scoring them, is clearly not feasible. There is, then, a search problem of finding the optimal interpretation—the one that *would* be chosen if all possible ones were generated and scored—without actually generating them all. My collaborator on the implementation of the program, Daniel Sleator, has devised a search procedure to solve this problem, which I now explain.

Our search procedure depends on the following idea. The score for an interpretation of a segment depends on the current COG; and this depends on the previous context. But all that matters to that segment is the COG of what precedes it. This means that if there are multiple interpretations of the prior context that all yield the same COG, only one of those—the highest scoring one—need be retained; there is no reason why the others would ever be preferred. The algorithm thus processes the piece in a "left-to-right" manner. At each point in the piece, it maintains a number of total interpretations; for each possible COG, it retains the highest scoring interpretation resulting in that COG. In considering a new segment, it tries all possible roots combined with each of the COGs.[15] This provides a number of continuations of the previous interpretations, each of them with a new COG and a new total score. But again, if several of these continuations result in the same new COG, only the highest-scoring one needs to be considered in analyzing the next segment. In this way we contain the explosion of possible interpretations, while still making sure that the highest-scoring interpretation is found. When the end of the piece is reached, the highest-scoring interpretation at that final segment is the optimal one for the piece overall.

As it has been described, the algorithm's only goal is to produce an optimal interpretation of the entire piece; it does not produce this result until the entire piece has been heard. This is clearly unsatisfactory as a model of listening; in listening, we are continuously processing the piece as we hear it, gradually building up our interpretation as we go along. In fact, however, the algorithm is extremely well-suited to capturing this aspect of perception. As described earlier, the algorithm is essentially processing the piece in a "left-to-right" fashion, one segment at a time. It does not commit to an interpretation of any part of the piece until the very end. But at any moment, there is one interpretation that is the highest-scoring one so far; this is the algorithm's preferred interpretation of the portion of the piece that has been heard. This brings us to another important feature of the algorithm. For each new segment it receives, the algorithm generates a number of new interpretations of the piece so far, again choosing the highest-scoring one. But suppose the highest-scoring interpretation at segment S_n involves a root at S_{n-1} other than the one originally chosen. Then, in effect, the algorithm is backtracking, revising its original interpretation of earlier events. Consider the altered version of measure 16, shown in Figure 15. When first heard (following the earlier context), the first half of this measure might well be analyzed as having root E; but once the following notes were heard, this might cause the algorithm to retroactively change its analysis of the first half of the measure, because this would improve the variance and strong-beat scores of the following segment, thereby improving the overall score. In this way, the algorithm could in principle capture the well-known phenomenon of retroactively revising something that is heard because of what happens afterwards. The ability of the algorithm to capture real-time effects such as this has not been tested, but it seems that the algorithm could offer a valuable model of real-time processing, in a way that

15. Because COGs are real numbers, not integers, some rounding off or "bucketing" of COGs is necessary. Similarly, because there are infinitely many roots, there must be some arbitrary limit on the range of roots considered (currently the algorithm considers four cycles of the line of fifths, or 48 steps).

does not depend on any additional procedures, but, rather, arises naturally out of the operation of the algorithm.[16]

Further Issues

First I will make some further points about the TPC representation. I will then discuss the issues of ambiguity and "grammaticality." Finally, I will consider the relevance of the current algorithm to the key-finding problem.

Little has been said about the operation of the TPC level. Recall that the main consideration here is that pitches should be spelled so that they are close together on the line of fifths. As with the harmonic representation, each pitch event in an interpretation is assigned a score, reflecting its closeness to the current COG; the total score for the interpretation is then the sum of all the individual pitch scores; the algorithm's goal is to find the interpretation with the minimum score. Like the harmonic representation, the preferred spelling of events depends on their context: the spelling of each event will depend on the spelling of previous events, and it might be desirable to respell events in light of subsequent events. Thus the algorithm must again consider total interpretations, choosing the interpretation that is optimal overall.

Earlier, we assumed that the TPC representation of a piece was complete when the harmonic representation was formed. In fact, however, the situation is more complicated. First of all, because the harmonic representation is continuously being built up in a left-to-right manner, it is necessary for the TPC representation to be generated in this way as well. One might still assume that each new segment of music was processed by the TPC level first and then by the harmonic level. But there is a further complication. The TPC level is, in a sense, prior to the harmonic level, because the harmonic level must have TPC information in order to evaluate different interpretations. But in many cases, it is desirable to have feedback from the harmonic level to the TPC level. A simple example of this is seen in Figure 16; consider the D♯ in the right hand. Simply in terms of TPC variance, D♯ is not strongly favored over E♭ here, if at all; they are roughly equal in closeness to previous TPCs. Yet D♯ is clearly the favored spelling. The current algorithm offers an explanation. If E♭ is chosen, then the TPCs present are E♭, F♯, and B, which do not form any tonal chord (that is, there is no

16. The ability of global preference rule systems to capture real-time effects has also been explored by Jackendoff (1991), who applies the preference rules of Lerdahl and Jackendoff's theory to real-time processing.

Fig. 16.

root for which they are all chord tones); if D♯ is chosen, however, a B-major chord is formed. (Remember that the compatibility rule is based on TPCs, not NPCs.) Thus the D♯ spelling is preferable to E♭ because it results in a better harmonic representation. But to capture such phenomena, there must be some mechanism for allowing harmonic factors to affect the TPC representation. The solution, in principle, is simple: search for the combined TPC-harmonic representation that receives the highest score overall, considering both the harmonic rules and the TPC variance rule. This is in fact the solution we adopt; however, our implementation of this is rather complex and will not be discussed here.[17]

As it stands, the algorithm outputs a single TPC and harmonic representation for the piece it is given. This is perhaps not ideal. Frequently ambiguities are present in the harmonic structure of a piece, and these ambiguities can be an important part of the piece's effect. In principle, the algorithm should be able to capture the "ambiguousness" of a passage well. If multiple interpretations of a passage are equally preferable, the algorithm should find several interpretations that are roughly equal in score. If there is clearly only one plausible interpretation, the algorithm should assign that interpretation a score much higher than all the others. The current system could also shed light on another important feature of pieces. What the algorithm outputs is not only a preferred interpretation but also a score for that interpretation. What this score indicates is how well the chosen interpretation satisfies the preference rules: to what extent it involves TPCs and roots that are close together on the line of fifths, reasonably few and short ornamental dissonances, and reasonably long chord-spans. If this score is low for the preferred interpretation, this means that *no* interpretation could be found that satisfied these criteria to a high degree. This could be used as a kind of measure of the grammaticality of the piece as tonal music. If one inputs, for example, a piece by (posttonal) Schönberg or Webern (or even Scriabin or neoclassical Stravinsky), I suspect that in general no high-scoring interpre-

17. Besides TPC variance and feedback from the harmonic level, one further consideration involved in the TPC representation is voice leading. A single pitch (such as A♭/G♯) may be spelled in different ways depending on its voice-leading context: for example, A-A♭-G versus G-G♯-A. The current implementation does not consider this factor, however. For further discussion, see Temperley (1996, pp. 186–194).

Fig. 17.

tation would be found for it. However, this idea should be approached with caution. It is not clear what "grammaticality" means when applied to music; grammaticality undoubtedly has other aspects besides harmony, and even harmonic grammaticality surely entails much more than the basic criteria discussed earlier. Scoring well on the algorithm might be offered as a necessary criterion for grammaticality, but it is surely not sufficient.

A final attractive feature of this algorithm relates to key structure. A considerable body of work has been done on the basis of key judgments. Krumhansl has proposed a model of key finding based on pitch distribution: the pitch distribution of a piece is matched to a "key profile," or characteristic distribution, for each key, and the profile with the closest match to the piece's pitch distribution is the chosen key (Krumhansl, 1990, pp. 77–106). However, Butler has pointed out that the same pitches arranged in different orders have different key implications (this has also been verified experimentally); this shows that pitch distribution alone cannot be sufficient to determine key (Butler, 1989, pp. 234–236). Other researchers, such as Winograd and Maxwell, have created algorithms that essentially use harmonic information as a basis for key judgments (Bharucha's connectionist algorithm is similar in this respect). However, it seems clear that harmonic structure alone is also not sufficient for key determination. Certainly it is easy to create passages with the same root progressions that imply different keys (C-Dm-G_7 implies C major, C-D_7-G implies G major). It is even possible to devise passages in which only the ornamental dissonances differ, which nevertheless have different key implications. Figure 17 gives an example; the first passage implies C major; the second, C minor. These demonstrations make it clear that neither pitch information alone nor harmonic information alone is enough to determine key; any robust key-finding algorithm must consider both.[18]

The current algorithm offers a promising basis for such a key-finding system. The algorithm outputs two representations of a piece: the TPC representation and the harmonic representation. The TPC representation provides a useful indicator of the pitch collection of a passage; this pitch

18. For a recent approach to key finding that uses both pitch and harmonic information, see Vos & Van Geenen (1996).

distribution could be summarized by finding the COG of the pitches in a passage, which would yield a single point on the line of fifths. The same could be done on the harmonic line of fifths; taking the COG of chord spans (weighting each one for its length) would provide a sort of harmonic "center" for the passage. Let us assume that each key has a characteristic "key point" on both the TPC line of fifths and the root line of fifths, representing the typical COG location for a passage in that key. (An evenly distributed C-major scale would have a TPC COG around D; with regard to harmony, the root COG in a typical C-major progression is probably between C and G.) The key of a passage would then be the one whose key points were closest to the COGs of the roots and TPCs in the passage. One useful thing about this approach is that it would distinguish between major and minor keys. Relative major and minor keys (such as C major and A minor) have similar pitch collections but different root progressions; thus they would be expected to have similar pitch COGs but different root COGs. Parallel major and minor keys (such as C major and C minor) have similar root progressions but different pitch collections; thus their root COGs would be similar but not their pitch COGs. In short, each key should be characterized by distinctive pair of pitch and root COGs. Preliminary tests have suggested that this approach could lead to quite a successful key-finding algorithm;[19] however, it has not been implemented computationally.

Results of the Implementation

I will now present results of two tests of the computer implementation of the algorithm (written by Daniel Sleator, using the language C++).[20] The program implements the algorithm exactly as it has been described. It accepts a piece represented as pitches in time (more precisely, as a list of notes with on times and off times). It also requires metrical information, in the form of several levels of equally spaced beats (we have developed a succinct way of encoding this). The program generates a TPC representation, mapping each pitch onto the line of fifths, and a harmonic representation, dividing the piece into chord spans and labeling each one with a root (which is again a point on the line of fifths). In doing this, the program searches for the combined TPC-harmonic representation of the entire piece that maxi-

19. See Temperley (1996, chap. 6) for examples and further discussion.

20. The program is publicly available; it can be obtained and used by anyone with access to a UNIX system. The program is available via the web site <http://www.cs.cmu.edu/~sleator/harmonic-analysis>. This web site contains all files necessary to run the program, as well as a number of input files, and a file README that explains how to use the program. The program accepts MIDI files as input; however, metrical information must be added by the user. Further explanation is provided at the web site.

mally satisfies both the harmonic rules and the pitch variance rule. Its primary output is a list of the notes, giving their TPC labels, and a list of chord spans, giving the time points and roots of each one; both levels are displayed in a graphic format similar to Figures 12 and 13.

Using the program, we have tested the algorithm on a number of pieces. On balance, we are pleased with the results; however, the tests have also pointed out several problems with the algorithm. Figures 18 and 19 gives two examples of the program's output, which are roughly representative of its level of performance: the unaccompanied melody "Yankee Doodle" and the Gavotte from Bach's French Suite no. 5 (discussed earlier). In each case, the program's TPC representation was perfectly correct. (In the case of "Yankee Doodle," the correct TPC representation seems obvious; in the Bach, it is given by the score.) At issue, then, is the harmonic representation. The harmonic representations generated by the program are represented in Figures 18 and 19 with root names above the staff; each root name indicates a chord span that begins on the onset of the note beneath and continues to the beginning of the next span. Where I consider the algorithm's choice to be definitely incorrect, I have included my own analysis in brackets. (In several other cases, the program's analysis differs slightly from my own, but seems plausible; these are left as they are.)

In the case of "Yankee Doodle," the algorithm's analysis is highly successful. There are two questionable choices: the second half of measure 3 (I would prefer a root of C over G here) and the first half of measure 6 (I would prefer G over D). But harmonizing the melody in this way would not be unreasonable. Unaccompanied melodies are an important test of the algorithm, because many of the harmonies are implied rather than fully stated. This melody provides a nice illustration of the algorithm's rules. The variance rule plays an important role, for example in the F♯ of measure 2; by the compatibility rule, F♯ would be the preferred root for this pitch, but the variance rule overrules it, preferring D as a root, because D is closer

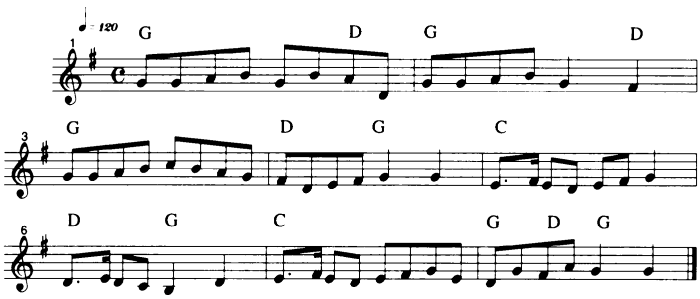

Fig. 18. The melody "Yankee Doodle," showing the algorithm's harmonic analysis. Each letter above the staff indicates a chord span of that root, beginning at the onset of the note beneath the letter and extending to the beginning of the next span.

to G on the line of fifths. The same applies to measure 5; here, C is preferred, although the compatibility rule would prefer a root of E. Although the compatibility rule is sometimes overruled, it is, of course, crucial; without it, the analysis would not in any way be constrained by the pitches in the piece. The strong-beat rule is also essential; while the fourth quarter of measure 1 (A-D) is given its own span, the eighth-note A in the first half of the measure is not, because this would mean starting the following span on a very weak beat; thus this note is treated as ornamental. Although many of the chord tones here could, in principle, be treated as ornamental (the second A in measure 1; the F♯ in measure 8), the penalty for this is high enough (because the time interval to the next stepwise event is fairly long) that another solution is preferred.

Figure 19 shows a second example, the Bach Gavotte discussed earlier. In the first half of the piece, the program's analysis is mostly correct. One oddity is the eighth-note D span in measure 6 (there are several eighth-note spans in the second half as well). In general, the algorithm prefers to avoid eighth-note spans at this tempo because of the strong-beat rule; this overrules any penalties that would result from treating a note as ornamental. However, when several ornamental notes are involved, their penalties sum together, making it preferable to treat them as chord tones. To my mind, there is a good deal of psychological reality to this; treating the fourth eighth of measure 6 as a chord is not implausible. This passage has one definite mistake: the algorithm treats the whole of measure 8 as one D span, while in fact the fourth quarter is clearly an A chord. The problem here is one that we have already discussed: the program is unable to handle escape tones, such as the final F♯ of the measure. Because the F♯ is not closely followed stepwise, it is not regarded as a possible ornamental dissonance, and the program must find some way of treating it as a chord tone.

The second half includes several problematic passages. The first is measure 12. The program assigns a root of D to the second half of the measure; the correct choice is F♯m♭5 (ii°/Em), a diminished triad. The problem here is that the algorithm has no notion of diminished triads ("♭5" is not among its set of acceptable TPC-root relationships); this is clearly a serious failing, which I hope to address. The analysis of the first quarter of the measure is also odd; C seems to be the correct choice. This points up another recurring problem with the algorithm: it often analyzes notes as ♭7s (as in the seventh of a dominant seventh) in cases in which this is not appropriate, such as the D in the left hand here. (Perhaps context could be used in some way to constrain the identification of ♭7s.) Second, the analysis of measures 18–20 is completely wrong. The problem here is that the passage consists largely of short notes in stepwise motion; in such cases, almost any note could be ornamental with a low penalty, so the algorithm

Fig. 19. The Gavotte from Bach's French Suite no. 5, showing the algorithm's harmonic analysis.

has difficulty choosing the right roots. One possible solution would be to make the algorithm prefer ornamental dissonances on weak beats, other things being equal. It can be seen, in this passage, that the first eighths of each half-measure are generally chord tones; perhaps the algorithm should give them more weight as determinants of the root (i.e., a higher penalty should be imposed for treating them as ornamental). This would also address the problem in measure 3 of "Yankee Doodle," noted earlier. A final problem here is the double neighbors in measures 16 and 19. Although in principle the algorithm can handle double neighbors, it seems that the ornamental dissonance penalty in these cases is slightly too high; in both cases, they are treated as chord tones instead (mistakenly, in my view).

One general limitation of the algorithm—mentioned earlier—is that it only labels chord segments with roots, omitting other information about chords such as mode and extension. For example, it does not distinguish between C major, C minor, C-dominant seventh, and C-minor seventh; all are simply labeled C. However, in the process of choosing a root for a chord, the algorithm must determine the pitch-root relationships of each pitch in the chord, and in many cases this could easily be used to find mode and extension information. For example, in choosing the root for the first chord span of the Bach Gavotte (measure 1), the algorithm must identify the Gs as 1/G, the Bs as 3/G, and the Ds as 5/G. Given that the span contains the pitch-root relationships 1, 3, and 5, and not b7, it follows automatically that the chord must be a major triad. Problems arise, however, in cases in which the notes of the chord are not fully stated. In measure 8 of "Yankee Doodle," the chord implied is presumably G major, but no 3 or 5 is present to indicate this. (On the other hand, in many cases, the exact identity of a chord seems to be psychologically indeterminate as well. In the second half of measure 7 of "Yankee Doodle," the root is clearly D; but is the implied chord D major or D_7?) Until this problem is solved, the algorithm's performance remains somewhat incomplete.

It is difficult to assess the program's performance overall—in comparison, for example, with the earlier programs of Maxwell and Winograd. It is not clear that its performance on the Bach Gavotte is superior to that of Maxwell's program on a comparable piece, shown in Figure 2. However, several points should be made in favor of the current algorithm. As discussed earlier, several phenomena, such as arpeggiations and implied harmonies, that are absolutely ubiquitous in tonal music are generally well handled by the present algorithm ("Yankee Doodle" is a case in point), but not by the earlier programs. Second, the earlier programs made use of information such as key signatures, spellings, and rhythmic notation, which is not available in listening; the current algorithm does not. The current algorithm is also (at a conceptual level) vastly more simple and parsimonious than the earlier programs, requiring only 5 rules, in contrast to the 36 rules of Maxwell's algorithm. (Even the problems with the algorithm discussed earlier seem to be fairly general ones, which might be solved with the addition of a few further rules of a similar character, rather than requiring many ad hoc adjustments.) A final issue concerns acquisition: how might a preference rule system such as this one be learned? Both Bharucha and Parncutt offer explanations for how their models might be acquired, and this is certainly a point in their favor. But before the acquisition question can be addressed, one must have a model that is reasonably adequate for the task at hand. (The learning issue is another reason that parsimony is important; from a developmental perspective, a model with only a few rules seems more plausible than one with many.) I have no answer to this question at present, however, and ultimately it will have to be addressed.

References

Bharucha, J. J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology, 16,* 485–518.

Bharucha, J. J. (1987a). MUSACT: A connectionist model of musical harmony. Proceedings of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.

Bharucha, J. J. (1987b). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception, 5,* 1–30.

Bharucha, J. J. (1991). Pitch, harmony and neural nets: A psychological perspective. In P. M. Todd & D. G. Loy (Eds.), *Music and connectionism.* Cambridge, MA: MIT Press.

Butler, D. (1989). Describing the perception of tonality in music: A critique of the tonal hierarchy theory and a proposal for a theory of intervallic rivalry. *Music Perception, 6,* 219–242.

Cuddy, L. L., & Badertscher, B. (1987). Recovery of the tonal hierarchy: Comparisons across age and levels of musical experience. *Perception and Psychophysics , 41,* 609–620.

Dennett, D. C. (1978). *Brainstorms.* Cambridge, MA: MIT Press.

Deutsch, D. (1982). The processing of pitch combinations. In D. Deutsch (Ed.), *The psychology of music.* New York: Academic Press.

Gjerdingen, R. O. (1987). *A classic turn of phrase: Music and the psychology of convention.* Philadelphia: University of Pennsylvania Press.

Jackendoff, R. (1991). Musical parsing and musical affect. *Music Perception, 9,* 199–230.

Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Perception, 8,* 189–202.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch.* New York: Oxford University Press.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review, 89,* 334–368.

Laden, B., & Keefe, D. H. (1991). The representation of pitch in a neural net model of chord classification. In P. M. Todd & D. G. Loy (Eds.), *Music and connectionism.* Cambridge, MA: MIT Press.

Lerdahl, F. (1988). Tonal pitch space. *Music Perception, 5,* 315–349.

Lerdahl, F. (1992). Pitch-space journeys in two Chopin preludes. In M. R. Jones & S. Holleran (Eds.), *Cognitive bases of musical communication.* Washington, DC: American Psychological Association.

Lerdahl, F. (1996). Calculating tonal tension. *Music Perception, 13,* 319–363.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

Longuet-Higgins, H. C. (1962). Two letters to a musical friend. *The Music Review, 23,* 244–248, 271–280.

Marr, D. (1982). *Vision.* San Francisco: W. H. Freeman.

Maxwell, H. J. (1992). An expert system for harmonic analysis of tonal music. In M. Balaban, K. Ebcioglu, & O. Laske (Eds.), *Understanding music with AI.* Cambridge, MA: MIT Press.

Narmour, E. (1990).*The analysis and cognition of basic melodic structures: The implication-realization model.* Chicago: University of Chicago Press.

Narmour, E. (1992).*The analysis and cognition of melodic complexity.* Chicago: University of Chicago Press.

Palmer, C., & Krumhansl, C. L. (1987). Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. *Perception & Psychophysics, 41,* 505–518.

Parncutt, R. (1989).*Harmony: A psychoacoustical approach.* Berlin: Springer-Verlag.

Pylyshyn, Z. W. (1989). Computing in cognitive science. In M. I. Posner (Ed.), *Foundations of cognitive science.* Cambridge, MA: MIT Press.

Scarborough, D. L., Miller, B. O., & Jones, J. A. (1991). Connectionist models for tonal analysis. In P. M. Todd & D. G. Loy, (Eds.), *Music and connectionism.* Cambridge, MA: MIT Press.

Schmuckler, M. A. (1989). Expectation in music: investigation of melodic and harmonic processes. *Music Perception, 7,* 109–149.

Shepard, R. N. (1982). Structural representations of musical pitch. In D. Deutsch (Ed.), *The psychology of music.* New York: Academic Press.

Tan, N., Aiello, R., & Bever, T. G. (1985). Harmonic structure as a determinant of melodic organization. *Memory & Cognition, 9,* 533–539.

Tanguiane, A. S. (1994). A principle of correlativity of perception and its application to music recognition. *Music Perception, 11,* 463–502.

Temperley, D. (1996). *The perception of harmony and tonality: An algorithmic approach.* Unpublished doctoral dissertation, Columbia University, New York.

Terhardt, E. (1974). Pitch, consonance and harmony. *Journal of the Acoustical Society of America, 55,* 1061–1069.

Vos, P. G., & Van Geenen, E. W. (1996). A parallel-processing key-finding model. *Music Perception, 14,* 185–224.

Winograd, T. (1968). Linguistics and the computer analysis of tonal harmony. *Journal of Music Theory, 12,* 2–49.

Zatorre, R. J. (1984). Musical perception and cerebral function: A critical review. *Music Perception, 2,* 196–221.