THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

On Interpreting Bach

Citation for published version:
Longuet-Higgins, HC & Steedman, M 1987, On Interpreting Bach. in Mental processes: Studies in cognitive science. MIT Press, pp. 221-241.

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Mental processes

OPEN ACCESS

15

On Interpreting Bach

H. C. Longuet-Higgins and M. J. Steedman
Department of Machine Intelligence and Perception
University of Edinburgh

Abstract

We have attempted to discover formal rules for transcribing into musical notation the fugue subjects of the Well-Tempered Clavier, as this might be done by an amanuensis listening to a 'dead-pan' performance on the keyboard.

In this endeavour two kinds of problem arise: what are the harmonic relations between the notes, and what are the metrical units into which they are grouped? The harmonic problem is that the number of keyboard semi-tones between two notes does not define their harmonic relation, and we further develop an earlier theory of such relations, arriving at an algorithm which assigns every fugue to the right key and correctly notates every accidental in its subject.

The metric problem is considered *de novo*, and a metrical algorithm is described whose failures to generate Bach's notation are as illuminating as its successes.

INTRODUCTION

The performance of a piece of music involves both the performer and the listener in a problem of interpretation. The performer must discern and express musical relationships which are not fully explicit in the musical score, and the listener must appreciate relationships which are not explicit in the performance. How the performer should convey his interpretation of the piece is an aesthetic question of the utmost delicacy; but the converse process, that of listening to a piece and discerning its structure, is partly amenable to objective investigation. This is because European classical music is written in a notation which conveys to the performer a considerable amount of information about its structure, and this information can be reconstituted by the educated listener from even a mediocre performance. We refer particularly to the time signature and the key signature, of which the former indicates the metric grouping of the notes and the latter their

harmonic relations with one another and in particular with the keynote. Any music student who scored the National Anthem thus:
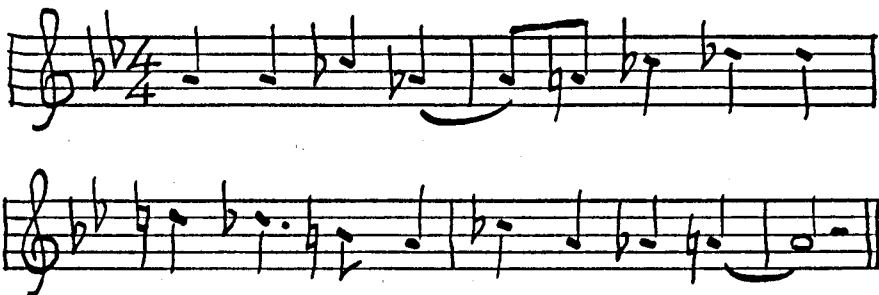
would get very low marks for dictation, even though the note lengths, and their positions on the keyboard, are correctly indicated. The 'correct' annotation of the melody in question is, of course:
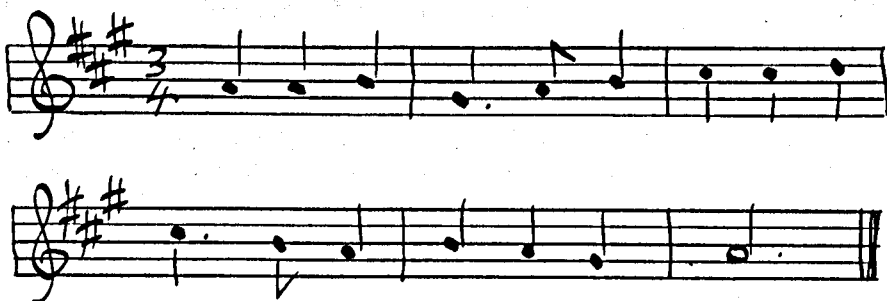
and even if he had never heard this melody before a competent musician would realize that this is how it ought to be scored.

As far as we are aware, no musical theorist has hitherto formulated the rules which generate the 'correct' score of a simple melody, as opposed to any of the numerous incorrect scores which provide the same explicit information. This gap in musical theory is all the more glaring in view of the considerable effort which has been devoted to much more ambitious undertakings, such as programmed musical composition. We are cynical enough to believe that it is only the prevailing babel in contemporary classical music which saves most of these compositions from being treated with the derision which they merit, and that if any progress is to be made in this direction it will first be essential to formalize the most elementary facts about musical competence, such as those we have just mentioned.

The work which we describe in this paper may suggest a comparison with recent work in linguistics, and in some ways the comparison would be apposite. The close connection between music and language is evident in song, where the rhythm of the words is mirrored in the rhythm of the melody to which they are set, and the rise and fall of the melody is to some extent

constrained by the intonation pattern of the words. But to say this is only to prepare the ground for a linguistic approach to classical music. The main point of comparison between our work and modern grammatical theory is that we attempt to make a formally precise model of the cognitive processes involved in the comprehension of classical melodies – in particular the fugue subjects of *Das Wohltemperierte Klavier*. Our choice of these particular melodies was to some extent arbitrary, but we felt that the choice justified itself by the great variety, rhythmic subtlety, and harmonic sophistication of the 48 fugue subjects and the universal admiration in which they and their composer are held by educated musicians.

We might have attempted to formulate our model as a generative grammar, but felt that such an attempt – even if it had been suited to the problem in hand – was clearly doomed to failure. It was unsuitable because those elements of structure in which we were interested, namely, the more obvious metrical and harmonic relationships, were relatively superficial and were likely to be largely independent of the underlying architecture. In this respect our analysis was 'phonological' rather than 'syntactic', and was, furthermore, designed to describe the mental processes of the listener rather than the composer. A generative grammatical approach could hardly succeed, we felt, unless it were founded on musical insights as deep as, and more explicit than, those of Bach himself. Such an approach might be feasible for nursery tunes, or even for Anglican chants, but for the Forty-Eight it would be a monstrous impertinence.

What we have in fact done is to write two 'parsing' programs, one for determining the metre and the other for explicating the harmonic relations between the notes of a Bach fugue subject. In writing these programs, which take account only of the note lengths and positions on the keyboard, we have attempted to make explicit our intuitive understanding of musical rhythm and harmony in general, and also to take account of one or two stylistic features which seem to distinguish Bach from some other classical masters. We do not claim that our parsing rules, or their discovery, have been informed by any methodological principles. Doubtless there are such principles, but we feel that they are more likely to become apparent after, rather than before, the formulation of rules to which they might apply. We have, however, attempted to make our programs mirror the progressive character of musical comprehension – by which we mean that as a fugue subject proceeds the listener's ideas about its metre and key become more and more definite, and may indeed crystallize well before the end of a long subject. The progressive nature of the listener's comprehension is made explicit in an assumption about the permitted order of musical events in an acceptable melody. This assumption we call the 'rule of congruence', and it is fundamental to the operation of both our harmonic and our metrical rules.

Both the key signature and the time signature of a melody indicate underlying structures with which individual notes or musical features may not be

superficially consistent. When some attribute of a note, such as its pitch, is consistent with the relevant underlying structure, we may describe the note as 'congruent' in this attribute, and otherwise as 'non-congruent'. Thus a note which does not belong to the scale defined by the key signature – an 'accidental' – is harmonically non-congruent; and a note carrying a rhythmic stress imparted by the context but not indicated by the time signature – a 'syncopated' note – is metrically non-congruent. (In adopting this definition of congruence we follow that introduced by Cooper and Meyer (1960) in the domain of rhythm, and extend it to harmony. We also follow their terminology in our use of the terms 'rhythm' and 'metre'.)

If our rules are to be able to decide the time signature and the key signature from the durations of the notes and their positions on the keyboard, some assumption must be made about how much of these data may safely be assumed congruent, and may therefore be used as evidence in reaching the required decision. The rule of congruence is such an assumption. It states that until a metric or harmonic signature has been established, every note will be congruent with it in the relevant attribute, unless the note is non-congruent with all possible signatures. In other words, a non-congruence must not occur until it can be recognized as such. This is surely common sense. Music would be a dull affair if all notes had to be in the key and all accents on the beat, but it would be incomprehensible if the key and metre were called into question before they were established.

The rule of congruence will be put to detailed use in the sections that follow; but we should say at once that it will turn out to be an oversimplification, albeit a fairly successful one. Most of our rules are, furthermore, of reasonable generality, and we expect them to have validity outside the particular corpus which we have studied – and indeed to the music of other composers, though this remains to be seen.

The metrical program, which we describe first, is less sophisticated than the harmonic program, and leaves a great deal of room for further work, though the results of its application are not without interest. Its principal limitation is that it attends only to the relative durations of the notes (and rests), but not at all to their positions on the keyboard, so that it is unable to yield any significant information when applied to those fugue subjects in which all the notes are of equal length. In such melodies the metrical structure must be inferred from the harmonic information, and as yet we have no specific ideas as to how this can be done, though a good musician can do it without the slightest difficulty.

The harmonic program, which has to decide as early as possible on the key and thereafter to place each note in its correct relation to the keynote, is further advanced. It is based on some earlier work by one of us (Longuet-Higgins 1962), in which the tonal relationships of classical music were exhibited as relations in a three-dimensional lattice, defined by the octave, the fifth, and the major third. This program attends only to the position

of each note on the keyboard (identifying all notes separated by octaves) and disregards their durations. To disregard the note lengths in a harmonic analysis might seem as high-handed as to disregard their values in a metrical analysis; but the results strongly suggest otherwise.

Finally, before embarking on the details, we should perhaps apologize for offering such a chapter to a volume of papers on Machine Intelligence. There is nothing in our programs which does more than express our own provisional views as to how an educated musician makes sense of a melody which is played to him, note by note, on the keyboard. The programs do not discern any rules for themselves, if only because we have no idea how to write a program capable of forming a description of a composer's musical style. We hope, however, that our work may be of interest as a case study in which the discovery of a suitable representation is found to be essential for the formalization of a cognitive process.

THE METRICAL ALGORITHM

The problem to which we address ourselves in this section is that of grouping the notes and rests of each fugue subject into metrical units. The longest type of metrical unit which is explicitly indicated in the score is the 'bar', and we would be very pleased to be able to place all the bar lines correctly on the basis of the notes of the subject alone. But even if we cannot identify whole bars we may hope to be able to identify the metrical units into which the bars are divided, or the smaller metrical units into which these in turn are subdivided; whether our identification is correct or not can be checked by inspecting the time signature which always appears at the beginning of the fugue, and indicates how the bar is to be metrically divided and subdivided. Thus the time signature 'c' or '¢' (short for 'common time') indicates that the bar is to be divided into two half-bars, which in turn are to be divided into two quarter-bars, and so on. But it is also possible for a metric unit to be subdivided into three lesser units, and when this happens it is indicated by a time signature such as 3/4, which means that the bar is divided into three crotchets (quarter-notes) – a situation which is to be distinguished from that signified by 6/8, which means that the bar is first divided in two, and that each half-bar is further subdivided into three quavers (eighth-notes). In Bach's time every metrical unit less than or equal to a bar was subdivided, if at all, into two or three smaller units, but never, for example, five or seven; we have taken this for granted in our metrical analysis.

Before describing the analysis, we shall say a word or two about notation. The note durations which to an American musician are known as a whole-note, a half-note, a quarter-note, an eighth-note, a sixteenth-note, and a thirty-second note are to an English musician a semibreve, a minim, a crotchet, a quaver, a semiquaver, and a demi-semiquaver respectively; to increment any of these durations by one-half, one writes a dot after the symbol representing the note (or rest). As for time signatures, we use in

table 1 a rather more explicit notation than the conventional one. Thus in table 1 the signature 3.2/8 (instead of 3/4) means that the bar is divided into three units each of which comprises two quavers (eighth-notes). An analysis yielding the signature 2/8 (with the right pairing of quavers) is then seen to be correct as far as it goes; but the result 2.2/8 would reveal a mistake at the second stage of grouping. If at any stage an analysis gives metrical units comprising the right number of smaller units, but incorrectly grouped together, then we indicate this fact by an asterisk against the corresponding number in the time signature. Thus the result 3*.2/8 means that the quavers were correctly grouped in pairs, but that the resulting crotchets were wrongly grouped in threes.

The rules which we develop below operate solely on the relative durations of the notes and rests, as they are given in the score. In an actual performance a player will 'phrase' the music by altering slightly the time values, to help the listener perceive the metrical structure; but our algorithm is designed to operate on a 'dead-pan' performance, in which even the pitch of each note is disregarded. Its limitations, which are quite severe, and the occasional wrong answer which it yields, may perhaps provide clues as to what extra information the performer should provide in carrying out his side of the business of 'interpretation'.

As one listens to a fugue subject, such as that of Fugue 2 in Book I, one gains the impression of certain notes being 'accented'. An accented note is one which is felt to fall at the beginning of a metrical unit of some length, which may be equal to the length of the note, or longer, or shorter. The simplest situation is that in which a metre has already arisen in the listener's mind, and a note is sounded at the beginning of a metrical unit; then the note, however short it is, automatically acquires an accent heavier than that of any other note in the same metrical unit. An example of this is Fugue 7 in Book II, where the listener adopts the initial semibreve as the unit of metre, and therefore feels a 'semibreve accent' on the second note, the fourth note, and so on, though none of these notes is as long as a semibreve. This fugue subject illustrates, in fact, two general rules about the establishment of a metre. The first is that, whatever its length, the first note of a subject (or the first two notes, if the second is shorter than the first and third: *see* below) may always be taken to define a metrical unit at some level in the hierarchy – though usually at rather a low level. This may be seen as a manifestation of the rule of congruence, and would, for example, exclude the unlikely possibility that the first note of the subject occurred at the end of a bar and was tied over the bar line. The second rule is that once a metrical unit has been adopted it is never abandoned in favour of a shorter one, or another one which cuts across it. The only way, then, in which the metrical hierarchy can be built up is by the progressive grouping of metrical units into higher units. Any rules for doing this must clearly be framed with great circumspection, since any mistake, once made, will vitiate all that follows.

226

What, then, is good evidence for enlarging a metrical unit which has already been established? A natural suggestion would be that if a note, falling at the beginning of a metrical unit, lasts for two or three units, then the metrical unit should be doubled or trebled accordingly. This suggestion would account very neatly for our appreciation of the metre of Fugue 9 of Book I, which begins with a quaver, establishing a quaver metre, and continues with a crotchet, which lasts two quavers and therefore becomes the new metric unit. And stated thus cautiously, the suggestion does not, in Fugue 8 of Book I, call upon us to abandon the crotchet metre established by the first note in favour of a dotted crotchet metre, for which the unit would be only 1½ times as long. But it raises a serious problem in Fugue 2 of Book I. Here the semiquaver metre gives place to a quaver metre when the third note is reached, and no problem arises for a while. But eventually we come upon a crotchet, and the above suggestion would lead us to adopt this as establishing a crotchet metre which actually cuts across the bar lines in the score. In fact no musician would be so deceived; but how does he avoid deception, and come to realize that the crotchet is in fact a syncopation?

The answer, we believe, lies in the early occurrence, in this subject, of a rhythmic figure which seems to play a central role not only in Bach's music but in the music of every succeeding generation. We name this figure the 'dactyl', after its counterpart in the metre of poetry. It consists – in its simplest manifestation – of a note followed by two equal notes of half the length, these being followed by a longer note. With no exceptions in the Forty-Eight, we find that if the first note of such a dactyl occupies one unit of an already-established metre, then the metrical unit must be doubled so as to accommodate the dactyl. Thus in Fugue 2 of Book I notes 5, 6, and 7 form a dactyl, establishing a crotchet metre into which the dactyl 10, 11, 12 fits neatly; there is another dactyl 14, 15, 16 soon afterwards, which cuts across the crotchet metre, but is powerless to overthrow it.

There are, however, other sorts of dactyl to which we undoubtedly pay attention in discerning metre. In the first fugue of Book I, for example, we find a pair of demi-semiquavers preceded by a dotted quaver and followed by a quaver. It is difficult to resist the view that the dotted quaver and the two demi-semiquavers provide the same metrical information as would a quaver and two semiquavers, so we count this figure as a dactyl too. As a result, the initially-established quaver metre is doubled so as to accommodate the dactyl, and gives way to a crotchet metre.

We may define a dactyl in general terms as the first three notes in a sequence of four, such that the second and third are equal in length and shorter than the first or the fourth. But we must then be very careful how we state the metrical implications of its occurrence. Consider, for example, Fugue 14 of Book I. The first two notes establish a crotchet metre, and are followed by a dactyl consisting of a semibreve and two quavers (followed by a minim). The dactyl is five crotchets long, but we are not allowed to group metric units into

227

fives. At this point we could of course draw in our horns and say that only the figures described earlier count as dactyls, but we prefer to restrict the range of the dactyl rule by a clause which says that the general dactyl under consideration is only to be adopted as the new metrical unit if it occupies a 'reasonable' number of current metric units, a 'reasonable' number being an integer whose only prime factors are 2 or 3. We shall see later what can be inferred when this condition is not satisfied.

Earlier we suggested that if a note, falling at the beginning of a metrical unit, lasted for two or three such units, then the metrical unit should be doubled or trebled accordingly. The natural generalization of this would be to combine the metrical units into sets of 2, 3, 4, 6, 8, 9, 12, 16, and so on if any note lasts for such a number of units. But then we run into immediate trouble with Fugue 5 in Book I. After a flourish of eight demi-semiquavers we come upon a dotted quaver, of six times the length. If we take the dotted quaver as the new metrical unit, we make a mistake which is apparent as soon as the following semiquaver is succeeded by another dotted quaver. The root of the trouble, clearly, is that we did not wait long enough to notice that the semiquaver following the dotted crotchet was an isolated short note. For this particular fugue subject we should obtain a correct result if we treated the semiquaver and the preceding dotted quaver as a dactyl, occupying eight demi-semiquaver units, and thereby establishing a new crotchet metre. But this expedient would lead to trouble in other connections – for example, in Fugue 8 of Book I. Perhaps erring on the side of caution we therefore rule that when a note falling at the beginning of a metrical unit is followed by a single shorter note (which is followed by a longer note), the metrical unit is to be doubled, trebled, etc., only if the length of the first note *minus* that of the second is a reasonable number of current metric units. Then in Fugue 5 of Book I the demi-semiquaver unit is multiplied by 4 rather than by 6 on the evidence of the dotted quaver-semiquaver pair.

At this point it is appropriate to return to the question of what conclusions can be drawn from the occurrence of an 'unreasonable' dactyl. We propose the supplementary rule that when a dactyl occupies an 'unreasonable' number of current metrical units, then the metrical unit is to be doubled, trebled, etc., only if the length of the first note *minus* the combined length of the two short notes is a reasonable number of current metric units. If we re-examine Fugue 14 of Book I we find that this rule leads to a correct inference – the metric signature that it finds is 3/4 consistent with the actual time signature 2.3/4.

We are now in a position to state more precisely the conditions under which a new metre can be inferred from the lengths of the incoming notes and rests. If a note is the first note of the subject, or falls at the beginning of a unit of the most recently established metre, then various possibilities must be explored. (1) The note may be the first note of a dactyl – a fact which can only be established by examining the durations of the following three notes.

If the total length of the dactyl is a reasonable number of current metric units, then this must be adopted as the new metric unit. If not, but the length of the first note minus the combined length of the shorter notes satisfies this condition, then this difference in length is taken as the new metric unit (as in Fugue 14 of Book I). Otherwise (but there are no such cases in the Forty-Eight) things are left as they are. (2) The note may be followed by a single shorter note (followed in turn by a longer note) – a fact which can be established only by examining the following two notes. If so, we subtract the length of the short note from the end of its predecessor. If the result is a reasonable number of current metric units, it is adopted as the new metric unit; otherwise (but again this does not occur) the previous metre is maintained. (3) The note may be of neither of the above types, but may endure for 2 or more current metric units. If so, these units are combined together in groups of n, where n is the largest reasonable number of units which are filled by the note. (4) If the note is not of types (1), (2), or (3), then the current metre is retained.

The above procedure enables us to take a step upwards in the metrical hierarchy whenever we encounter an accented note or dactyl of sufficient length, an accented note being one which falls at the beginning of a unit of the current metre. But when applied to a subject such as that of Fugue 15, Book I, it fails to indicate the triple grouping of the quavers in the first bar – though this is quite obvious to the hearer – so that on reaching the crotchet in the second bar it incorrectly suggests a crotchet grouping of the quavers. To deal with such cases we need to extend the concept of an accent to metrical units as well as to individual notes. A metrical unit is 'marked for accent' if a note or dactyl begins at the beginning of it and lasts throughout it. We can now state a new rule for establishing metre at a higher level, as follows: if, in the current metre, a unit which is marked for accent is followed by a number of unmarked units, and then by another marked unit, which in turn is followed by an unmarked unit, then the two marked units are taken to establish a higher metre, in which they occur on successive accents. In Fugue 15 of Book I the first two quavers of the subject are, according to this definition, both marked for accent, and the second is flanked by unmarked quaver units, as required for the application of the rule, so the rule establishes a higher metre, of dotted crotchets, in which the first and the fourth quaver units of the subject occur as the first two accents. But there is an important qualification to the rule, namely that if the higher metrical unit so defined is an unreasonable multiple of the lower one, it must not be adopted. With this qualification the new rule, which we call the 'isolated accent rule', can be applied without absurdity.

We may illustrate the application of these rules in relation to Fugue 2 of Book I. The first note establishes a semiquaver metre, to which the second note conforms. The third note converts this into a quaver metre, fitting the fourth note. The fifth note is the first of a dactyl, lasting a crotchet, establishing

229

a crotchet metre in which the dactyl is marked for accent. Notes 8 and 9 constitute an unmarked crotchet unit; notes 10, 11, and 12 constitute another dactyl, which is also marked for accent. Notes 13 and 14 make up an unmarked crotchet unit, so the isolated accent rule can be applied to the two marked crotchet units, revealing a minim metre, which is in fact the metre of Bach's half bars. The fact that 14 is the first note of a dactyl does not disturb this analysis, since it is unaccented in the crotchet metre which has been established by the time it is sounded.

Table 1. Metrical analysis. > means that whole bars are grouped together, = means that notes are all equal in length, * indicates an erroneous result.

	Book I			Book II		
	Correct	Found	Comment	Correct	Found	Comment
1	$2^3/8$	$2/8$		$2^3/16$	$2^2/16$	
2	$2^4/16$	$2^3/16$		$2^3/8$	$2/8$	
3	$2^3/8$	$2/8$		$2^3/8$	$1/8$	=
4	$1/1$	$2/1$	>	$3.2^2/16$	$1/16$	=
5	$2^5/32$	$2^3/32$		$2^3/8$	$2/8$	
6	$3.2/8$	$2/8$		$2^3.3/16$	$2^*.3/16$	*
7	$2^4/16$	$2^2/16$		$1/1$	$1/1$	
8	$2^2/4$	$1/4$		$2^3/8$	$2^*.2/8$	*
9	$2^3/8$	$2/8$		$2/1$	$1/1$	
10	$3.2^2/16$	$2/16$		$2^2.3/8$	$2.3/8$	
11	$3/8$	$3/8$		$2.3/16$	$3.2/16$	*
12	$2^2/4$	$2/4$		$2^2/8$	$2/8$	
13	$2^3/8$	$2/4$		$2^4/16$	$2^5/16$	>
14	$2.3/4$	$3/4$		$2^3/8$	$2^*.2^2/16$	*
15	$2.3/8$	$3/8$		$3.2/16$	$1/16$	=
16	$2^3/8$	$2/8$		$3/4$	$1/4$	
17	$2^3/8$	$2/8$		$2^3/8$	$2/8$	
18	$2^2/4$	$1/4$		$2.3/8$	$1/8$	=
19	$3^2/8$	$1/8$	=	$2^4/16$	$2^*.2/16$	*
20	$2^4/16$	$2^2/16$		$2^2/4$	$1/4$	
21	$3.2/8$	$12/8$	>	$3.2/8$	$1/8$	=
22	$2/2$	$1/2$		$3/2$	$3^*/2$	*
23	$2^3/8$	$2^2/8$		$2/2$	$1/2$	
24	$2^3/8$	$2^2/8$		$3/8$	$1/8$	

On the basis of these rules – which may seem complicated, but are un-doubtedly much less complicated than our perceptual processes – we have written a program for assigning metrical structures to the 48 fugue subjects; the results are given in table 1. By and large, the program avoids mistakes, at the cost of a rather incomplete analysis: for example, it is powerless to deal with any subject in which all the notes are of the same length, and where the musician would obviously rely on harmonic clues or skilful phrasing for discerning the metre. But there are some interesting mistakes. In Fugues 8, 14, and 19 in the second Book the program makes just the sort of mistake which a musician would probably make in performing the same task. In all three Fugues the rhythm departs from the underlying metre before the latter has been made explicit; by thus violating the rule of congruence Bach plainly intends to throw the listener off the track. There are three other subjects for which the program gives a wrong result. In Fugue 6 of Book II it makes a mistake which would probably not be made by a musician who paid attention to the pitches of the notes. In Fugue 22 of Book II the program places accents on the second and fifth minim unit rather than the first and fourth, and a musician might well make the same mistake. And in Fugue 11 of Book II it is misled by the semiquaver rests which to a musician are merely indications of phrasing and are equivalent to staccato markings of dotted quavers. In three Fugues (4 and 21 of Book I and 13 of Book II) the program actually carries its analysis beyond the bar; the resulting bar groupings seem quite acceptable, though they do not carry Bach's imprimatur. But one of these analyses (of Fugue 21, Book I) must be regarded as a fluke, because all the program does is to group quavers in twelves, without specifying whether the metrical structure is 2.2.3/8, or 2.3.2/8, or 3.2.2/8.

Our main reaction to these results was one of surprise that a program operating on so little information, and embodying such simple rules, could reveal so much metrical structure.

THE HARMONIC ALGORITHM

The problem for which our harmonic program is designed is to determine the key of a fugue from its subject, and, if there are any notes not belonging to the key, to determine their relation to those which do, and to the keynote in particular. Once this problem has been solved, it is a trivial matter to tran-scribe the solution into standard musical notation, so we shall eschew notational technicalities as far as we can in the following paragraphs.

Figure 1 depicts a short section of the keyboard, with each note lettered or numbered according to its position in the octave. The key of C major comprises the 'white notes' 0, 2, 4, 5, 7, 9, and 11, and contains the three 'major triads' (C E G), (G B D), and (F A C), that is, (0 4 7), (7 11 2), and (5 9 0). In each triad the first and third notes are separated by a 'perfect fifth', which is one of the three basic intervals of music. The other two basic intervals are the octave – but we are treating two notes an octave apart as
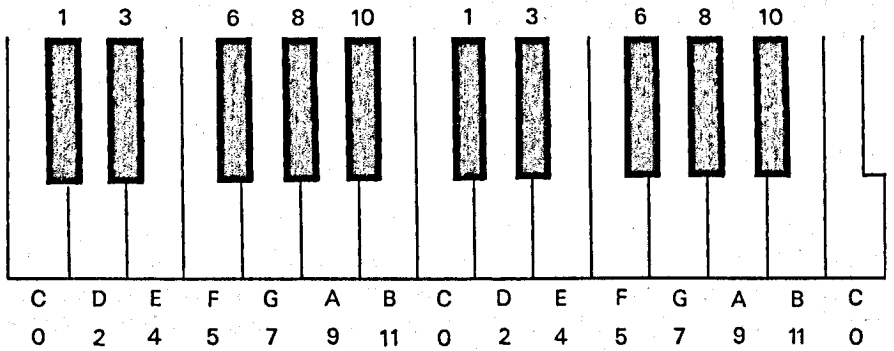
231

Figure 1

equivalent – and the 'major third', which is the interval between the first and second notes of each major triad. These facts enable us to represent the harmonic relations within C major by the two-dimensional array shown in figure 2, where each note is a perfect fifth below the note on its right and a major third below the note written above it. A major triad then forms an L-shaped cluster of three notes. One may set up a corresponding array of numbers, and extend it indefinitely in both dimensions (see figure 3) where a move to the right adds 7 modulo 12 and a move upwards adds 4.
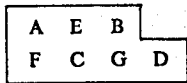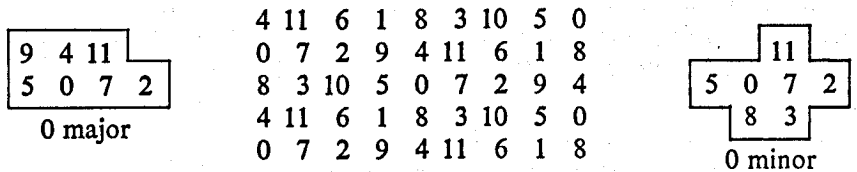

Figure 2

Figure 3

The key of 0 major then comprises the seven notes shown on the left. Notes other than these are conventionally denoted by 'sharps' ($\sharp$) and 'flats' ($\flat$), as in figure 4.
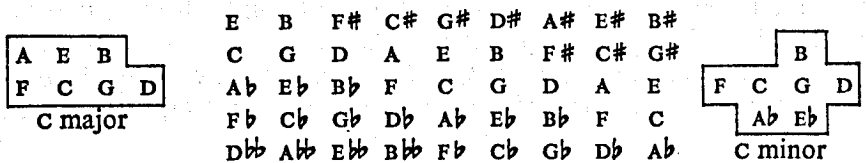

Figure 4

232

It is important to observe that notes which have the same number in figure 3 do not necessarily have the same symbol in figure 4. For example, a note which lies a major third above an E in figure 4 is called G♯, and a note which lies a major third below a C is called A♭, but both notes are played in position 8 on the keyboard. And even figure 4 blurs the distinction between, for example, the A which lies immediately above F and the A which lies four places to its right. Strictly speaking, all of the notes in figure 4 should have different symbols, because the octave, the perfect fifth, and the major third are incommensurable intervals. This incommensurability can be inferred from the fact that the frequency ratios of the intervals in question are 2/1, 3/2, and 5/4 respectively; but it is better to regard it as a musical rather than a mathematical fact, because the ear can be tricked by the keyboard into supposing that an octave equals three major thirds, and that two octaves plus a major third equals four perfect fifths. Indeed, it is just this kind of trickery which our harmonic program is designed to expose; given the notes of a fugue subject played on the keyboard it must decide how they should be related to one another on a two-dimensional lattice in which every point is properly distinguished from every other.

Armed with this two-dimensional representation of harmonic relationships we are now in a position to see how to assign a particular fugue subject to a particular key. A convenient illustration is provided by Fugue 9 of Book I of the Forty-Eight, where the notes of the subject include 1, 3, 4, 6, 8, 9, and 11, but not 0, 2, 5, 7, or 10. Consulting figure 3 we would notice that a major-key-shaped box could be fitted round the notes of the subject in just one way (see figure 5) and we would conclude that the key was 4 major – that is, E major. In fact there are rather few fugues in which Bach uses all, and only, the notes of a major key, and a more sophisticated analysis will usually be called for. But before embarking on this let us discuss minor keys, where the situation is rather less straightforward.
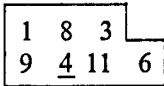

Figure 5

Like the major key, the minor key of Bach's tradition comprises seven principal notes. Thus the key of C minor is composed of the seven notes C, D, E♭, F, G, A♭, B, and C, the harmonic relations between them being fixed by the need to accommodate the major triad (G B D) and two 'minor triads' (C E♭ G) and (F A♭ C). In a minor triad the outer notes are a perfect fifth apart, but the middle note is a major third below the highest note, rather than a major third above the lowest. These facts can be represented by enclosing the notes of a minor key within a box of the shape indicated on the right in figures 3 and 4; in these figures the notes of C minor are seen to differ from those of C major by the substitution of A♭ and E♭ for A and E. But there is an additional

233

complication. In Bach's time, if a composer wished to write an ascending or a descending scale in a minor key he would use the *major* sixth in the ascending scale and the *minor* seventh in the descending scale. That is to say, in 0 minor (c minor) the upward scale 7, 8, 11 (G, A♭, B) would become 7, 9, 11 (G, A, B), and the downward scale 0, 11, 8 (c, B, A♭) would become 0, 10, 8 (c, B♭, A♭). This convention, which we shall call the 'melodic convention', means that 9 may be regarded as belonging to 0 minor only if it occurs in the context 7, 9, 11, and that 10 may be so regarded only in the context 0, 10, 8. As an example of the melodic convention we may consider Fugue 4 of Book II. The notes of the subject are, in order, 1, 0, 1, 3, 1, 3, 8, 10, 0, 1, 3, 4, 6, etc. We observe from figure 3 that there is a minor key, namely 1 minor, which includes all these notes except one (*see* figure 6). But the note 10, which is
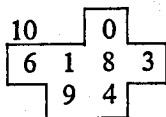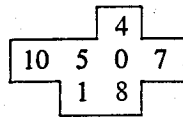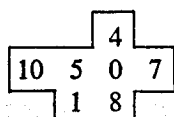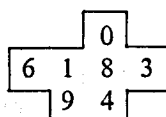

Figure 6

Figure 7

the major sixth of 1 minor, occurs within the scale 8, 10, 0, and so can be understood to arise from the melodic convention. This note therefore provides no evidence that the subject is not in 1 minor – actually it is.

It is by no means uncommon, however, for the notes of a fugue subject to wander right outside the original key, even when allowance is made for the melodic convention. As an example we may cite Fugue 12 of Book I, where the first few notes are 0, 1, 0, 11, 4, 5, etc. Figure 3 reveals that no key, major or minor, will accommodate a 0, a 1, and an 11. After 0, 1, 0, the note 11 excludes all possible keys, and must therefore be outside the initial key. But there is one key which includes all of the first six notes *except* 11, namely the key of 5 minor (*see* figure 7). This should therefore be the key of the fugue – as it is.

In some cases, then, it is possible to decide the key after hearing only a few notes of the subject. A striking example is Fugue 4 of Book I, where the subject comprises just four notes, namely 1, 0, 4, 3. There are two keys which include 1, 0, and 4, namely 5 minor and 1 minor (*see* figure 8); but of these

5 minor

1 minor

Figure 8

only 1 minor can also accommodate the fourth note. The key of the fugue is in fact 1 minor.

The above examples might suggest that the key of a fugue can always be determined by a process of elimination before the end of the subject; but this

is not so. A simple example is Fugue 9 of Book II, where the notes of the subject are 4, 6, 9, 8, 6, 4. At the end of the subject (that is, before the second 4, when another part enters) the key could be 4 major, or 9 major, or 1 minor (*see* figure 9).

To decide among these three possibilities we appeal to a musical intuition which is embodied in one of the standard rules of fugue. This rule demands that the first note of the subject be the tonic or the dominant (that is, the keynote or the note a perfect fifth above it) – though Bach actually breaks this rule in Fugue 21 of Book II. But to appeal directly to this rule of fugue would conflict with the spirit of our investigation, so we adopt instead a 'tonic-dominant preference rule' to the following effect: when a dilemma of choice presents itself, first preference is given to a key whose tonic is the first note of the subject, and second preference to a key whose dominant is the first note of the subject. A dilemma of choice may arise in either of two ways. First (as above) one may reach the end of the subject without having

1	8	3			6	1	8				0		
9	4	11	6		2	9	4	11		6	1	8	3
											9	4	

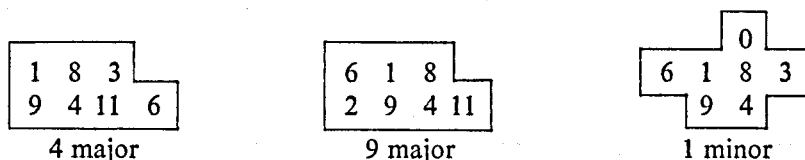4 major   9 major   1 minor

Figure 9

eliminated all keys except one; and, secondly, one may meet a note which does not fit into any key at all. The dilemma illustrated in figure 9 is of the former kind; here the first note (4) is the tonic of 4 major and the dominant of 9 major, but is neither the tonic nor the dominant of 1 minor, so that 4 major is selected as first preference. The latter type of dilemma is illustrated by Fugue 14 of Book I, where the subject begins with the notes 6, 8, 9, 8, 10, 11, 10, 8, 10, 0, 1, and so on. The first four notes are compatible with several different keys, including 6 minor, 4 major, 9 major, and 1 minor. But the fifth note, 10, excludes all these possibilities, so a dilemma arises. But until that moment 6 minor was a possibility, a key whose tonic is the first note of the subject, so 6 minor is correctly selected as the key of the fugue.

A program embodying these rules – and no others – assigns all 48 of the fugues to their correct keys, on the basis of the notes of the subject alone. (The last note of the subject is taken, for definiteness, to be the last note which is sounded before, or at the same moment as, the first note of the second entry.) For most of the fugues the key is decided well before the end of the subject; in only 17 out of 48 cases is the tonic-dominant preference rule appealed to at the end.

The task of key assignment is, however, only part of the harmonic problem. As already remarked, a subject may contain 'accidentals' – notes which lie outside the original key – and may indeed 'modulate' into a new key. A fine

example of this is the subject of Fugue 24 of Book I, shown in figure 10. The notes of the first bar correctly establish the key as 11 minor, but in the next bar there are three notes – 3, 0, and 5 – which do not belong to the key and are written as D♯, C♮, and E♯ respectively. In the third bar there are three more extraneous notes – 0, 9, and 8 – of which the 0 is now written as B♯. (The 9 and the 8 cannot be accommodated in 11 minor because they do not occur in the contexts demanded by the melodic convention.) The question arises: how can the listener tell that the six notes under consideration are to be interpreted as Bach wrote them, and in particular that the first 0 is a C♮ and the second a B♯?

Figure 10

For the first four of these notes the solution is provided by another rule, which seems to describe Bach's use of chromatic scales – note sequences of the form $(n, n+1, \ldots, n+m)$ or $(n, n-1, \ldots, n-m)$. We state this rule – the 'semitone rule' – as follows: in a chromatic scale the first two notes are always related by a diatonic semitone, and so are the last two. By a 'diatonic' semitone we mean the interval between B and C in C major – a move of one step to the left and one step downwards in a key diagram. Figure 10 shows that in the second bar of the above fugue subject there are three chromatic scales, namely $(4, 3)$, $(0, 11)$, and $(6, 5)$. But the notes 4, 11, and 6 all lie in the established key (11 minor – *see* figure 11). Therefore the 3, the 0, and the 5 are to be interpreted as D♯, C♮ and E♯ respectively. In the third bar the 0 is the last note of the chromatic scale $(2, 1, 0)$ and the first note of the scale $(0, 1)$. It must therefore be a diatonic semitone below 1 (which belongs to the key) and be interpreted (*see* figure 11) as B♯ rather than C♮. A problem arises, however, with the first 9 in the third bar. This lies outside 11 minor, but is not part of a chromatic scale. We therefore – and this is our last harmonic rule – place it in the closest possible relation to the notes which have already been heard. A convenient measure of closeness is the sum of

its city-block distances, in figure 11, from all the positions at which earlier notes have been placed, and this identifies it as the lower right-hand A in figure 11, which is undoubtedly correct musically, as it must be a major third below the preceding C♯. Finally the last note of the third bar has to be assigned a position in figure 11. If we disregard the trill, which involves the note a semitone above, and treat it as a plain unadorned 8, we find that there are two different readings of the note in figure 11 which give equal values for the sum of the city-block distances from already occupied positions. Either the left-hand or the right-hand G♯ in the figure would do equally well – though they would be annotated in the same way. This is the only case in which the city-block rule fails to settle the interpretation of an accidental – though it does indicate how it must be annotated. And even this single ambiguity would disappear if we took into account the upper note of the trill and demand that the two notes be inserted into figure 11 together at positions separated by a diatonic semitone.

```
   8   3 |10| 5   0          G# D#| A#| E# B#
 9 | 4  11  6   1 | 8       A | E   B   F# C# | G#
     | 0   7   2 | 9             | C   G   D | A
```

Figure 11

We may summarize the operation of the harmonic algorithm as follows. Each major or minor key is represented by a box of appropriate shape superimposed on a two-dimensional lattice like that in figure 3. The first note of the subject lies in 14 of these keys, but not in the other ten, which are immediately eliminated. As each note comes in, it is tested in relation to each surviving key to see whether it lies in that key; if not, the key is eliminated unless the context of the note permits it to be assigned to the key on the basis of the melodic convention. If any note is found to have eliminated all the 24 possible keys, then their elimination must be reconsidered. First it must be asked whether one of the keys just eliminated was a key of which the first note was the tonic; if so, then that must be taken as the key of the fugue. If not, did one of the keys just eliminated have the first note as its dominant? If so, then that must be the key of the fugue. If the answer to both these questions is in the negative, then all the keys just eliminated must be reinstated and the following note tested in the same way. If at any stage a note eliminates all keys except one, then this is taken to be the key of the fugue. And if at the end of the subject two or more keys have survived elimination, then the tonic-dominant preference rule is used to decide between them.

In this part of the algorithm the operation of the rule of congruence is apparent; until the key of the fugue has been identified, any note which does not fit into a given key must be counted as evidence against that key unless it happens to eliminate all possible keys – which often happens. But after the key has been established then accidentals can be met with equanimity

237

– provided that their relation to the established key can be clearly determined. The second part of the algorithm is designed for this purpose. It first invokes the semitone rule, which can be used for placing notes which belong to the first or the last pair in a chromatic scale, provided (as almost always happens) the other note of the pair belongs to the established key. If the semitone rule fails to place a note, then the algorithm resorts to the city-block distance rule for placing the note in the closest possible harmonic relation to the previous notes of the subject. In only one case does this rule also fail to place

Table 2. Key analysis (* tonic-dominant preference rule used).

number of fugue	key of fugue	Book I number of notes in subject	Book I decision at note number	Book II number of notes in subject	Book II decision at note number
1	0 major	16	16*	23	23*
2	0 minor	11	5	9	9*
3	1 major	22	16	4	4*
4	1 minor	5	4	22	12
5	2 major	15	15*	9	9*
6	2 minor	12	8	25	15
7	3 major	24	11*	20	20*
8	3 minor	12	12*	14	9
9	4 major	12	11	6	6*
10	4 minor	25	7*	46	18
11	5 major	19	6	21	17
12	5 minor	15	4*	27	7
13	6 major	17	8	28	12*
14	6 minor	18	5*	18	18*
15	7 major	31	15	43	16
16	7 minor	11	4	18	18*
17	8 major	7	7*	22	22*
18	8 minor	17	5	25	25*
19	9 major	7	7	20	20*
20	9 minor	31	5	10	5
21	10 major	38	14	24	9
22	10 minor	6	6*	27	5
23	11 major	14	11	12	12*
24	11 minor	22	7	25	6

a note unambiguously; the program is designed to draw attention to such failures, so we can be sure that this case is unique in the Forty-Eight.

Space would not permit us to quote all the fugue subjects, or to give our analysis of them in detail, so we conclude by simply appending the main results (*see* table 2). Suffice it to say that every key is assigned, and every accidental notated, in accordance with Bach's score; and, furthermore, that the placing of the notes on each key diagram, which indicates how the subject would have to be played in 'just intonation', rather than on the keyboard, seems to accord completely with our musical intuition. We know, however, that in at least one Bach fugue not belonging to the Forty-Eight, namely, the second Kyrie of the Mass in B minor, our program would assign the fugue to the wrong key because the second note of the subject violates the rule of congruence. But such liberties are the very stuff of artistic creation.

REFERENCES

Cooper, G. & Meyer, L. B. (1960) *The Rhythmic Structure of Music.* Chicago: Chicago University Press.

Longuet-Higgins, H. C. (1962) Letter to a Musical Friend; and Second Letter to a Musical Friend. *The Music Review*, **23**, August, 244–8; and November, 271–80.