The uses of space in music notation

John Sloboda

Abstract

Space is an essential notational dimension in music. The use of space in orthochronic notation (accepted today as the standard system) is described in some detail. The historical development of this use is outlined, with examples from earlier systems. The major uses of space are to provide means of notating pitch and duration, and to indicate the synchronisation of concurrent musical events. The question of how one might evaluate different ways of using space in music notation is then treated with reference to three levels of analysis: (1) the information about the music which needs to be represented, (2) the nature of the symbols (and their spatial characteristics) chosen to represent the information, and (3) the requirements of the reader.

1. Introduction

In this article I want to describe how space is used in music notation. Music notation is unfamiliar to many people and so I have attempted to make what I say intelligible to the non-musician by defining all the technical terms I use. Music notation (score) is very different from language notation (text) and so some prefatory comparative remarks may help to place this notation in a wider context.

Perhaps the most fundamental difference between score and text is that a score must be able to specify different events as occurring simultaneoulsy whereas text portrays a single sequence of events. The problem of how to link up parallel streams of information has thus been fundamental to the development of score. No analogous problem exists for text. A second difference concerns use. Score readers are mainly concerned with producing a musical performance. Text readers are more concerned with understanding and remembering what they read. This difference makes issues of layout of foremost importance in a score. The music reader cannot afford to lose his place or experience ambiguity even for a second if he is to maintain the flow of performance. This demand has expressed itself in intense and prolific experimentation with different ways of arranging score material over the centuries. In contrast, the solitary text reader is able to pace his own reading to accommodate deficiencies in layout. Thus there has

Visible Language, XV 1, pp. 86-110.
Author's address: Department of Psychology, University of Keele, Keele ST5 5BG, England.
0022-2224/81/0001-0086$02.00/0

not been the same historical incentive to experiment with layout. Unhelpful layout may be a nuisance: but it hardly ever leads to total breakdown. A third difference concerns the levels at which spacing and layout become important. Score material presents the reader with diverse spatial information at a microscopic level. Any two-centimeter square of a modern score will contain a rich array of symbols at varying distances and directions from one another. In contrast a similarly sized portion of text is most likely to contain evenly spaced letters in evenly spaced rows. Interesting spatial differentiations occur, if at all, at the macroscopic level, and concern such matters as how paragraphs are sited. Another way of putting this difference is to say that in a score there are complex spatial constraints which determine the positioning of each notational element with respect to its neighbors, whereas in text the determination of the position of one letter in relation to its neighbors is trivially simple. Thus, layout is an integral part of the music notation system. In language notation it is an optional extra.

There are, of course, many points one could make which would tend to blur the clear lines drawn up by these distinctions. There is not *one* use for scores any more than there is for texts. It would be fair, however, to claim that these differences do characterise broad and prevailing tendencies within the two systems.

What follows from these comparisons? Firstly, spatial considerations in music notation are more complex and multi-levelled (and arguably more interesting) than spatial considerations in text. Secondly, perhaps as a result of the complexity, there has been little, if any, controlled experimentation using different spatial arrangements of music. And so, thirdly, this article cannot draw on published psychological research. Rather its aim is to present an historical and conceptual perspective on some of the major motivations for, and problems with, using space as a notational tool. In this task I draw most heavily on the work of musicologists and music historians, in particular the monumental study of Apel (1953). Other useful source and background material has been provided by Abraham (1979), Cole (1974), Hyatt-King (1964), Karkoschka (1972) and Read (1974).

Western musical civilisation of the last 1500 years has produced a bewildering multiplicity of notational systems devised in different contexts and for different purposes. Nonetheless, today we would recognise one particular system as central. This is, to adopt Read's (1974) terminology, the *orthochronic* system. Figure 1 shows a typical portion of this notation. It has been the major western notational system for

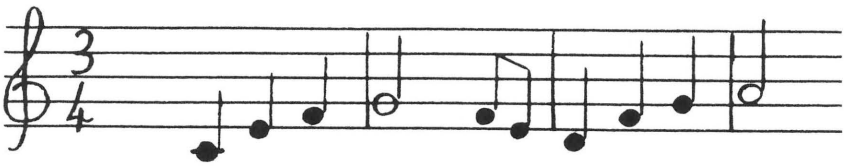

Figure 1 A section of orthochronic notation showing the principal notational features.

over 400 years, holding its place against all competitors even in this century which has seen a mushrooming growth of new notations. As the discussion proceeds I shall touch on various aspects of orthochronic notation, looking both at their development and at notational systems which have used space in different ways to achieve similar ends. Also, since any notational system can only meet limited needs, it is necessary to compare orthochronic notation with systems whose aims are different. Differing aims place differing demands on the available space.

2. Directions for action

A most important fact about orthochronic notation is, apparently, negative. It is impossible to tell, just by examining a typical extract, which instrument the extract is intended for. The central features of the notation are abstract. They specify pitch and rhythmic relationships between notes and groups of notes, *not* what keys are to be pressed on an instrument *nor* precisely how the notes are to sound. They are intended to convey something of the musical structure to a reader. Psychological implications of this fact have been discussed by Sloboda (1978a, in press). This characteristic probably explains why orthochronic notation has retained its central position. It is a *lingua franca* for all musicians, as understandable to a violinist as to a singer.

The corollary of this is that each instrumentalist must have additional practical knowledge before he can actually play the music. For instance, orthochronic notation may tell a flautist to play the note *D*, but it does not tell him which fingers to place over which holes to achieve the *D*. The wish to have a more direct notational aid to performance has sometimes led to the development of notations based upon a different principle – that of specifying actions on a particular instrument. Such notations achieve this end at the expense of universality. A notation devised for the violin would be meaningless to a trumpet player.

One large sub-group of such systems are the tablatures. Figure 2 is part of a French lute tablature of the sixteenth century. The horizontal lines represent the strings of the lute. The letters designate frets on the finger-board. In this example, the frets are marked by the letters B to H, with A denoting an open string. Tablatures use letters or numbers to specify at least part of the actions required. Some modern successors have eliminated alphabetic symbols completely in favour of graphic representation. A simple example of this is modern guitar notation (Figure 3) in which strings and frets are represented on a two-dimensional matrix, with dots to show the finger positions.

A more complex example is a modern system for the piano called Klavarscribo. In this system vertical lines represent the black keys on the piano, the spaces between representing white notes. The score is read from top to bottom, and the small circles represent the keys to be struck (unfilled circles are white notes). Figures 4 and 5 show the same extract from F. Chopin's Sonata op. 58 written in orthochronic notation and Klavarscribo. To the eye there is little to suggest that it is the same music in both cases.

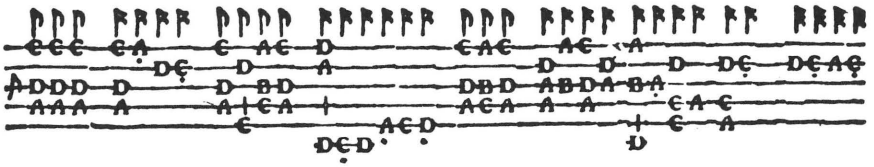

Figure 2 "Fortune a bien couru sur moi" from *Tres breve et familiere introduction pour entendre et apprendre par soy mesme a iouer toute chanson reduictes en la tabulature du Lutz*. Source: Apel (1953) p. 65.

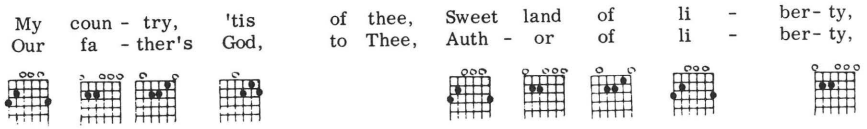

Figure 3 Chord sequence for "My Country, 'tis of thee" (first line) in modern guitar notation.

When we go on to consider the detailed requirements of a developed music notation system, I shall wish to return to these examples to argue that the price paid for instrumental specificity is, in general too high, since it restricts the number of dimensions remaining for representing other important aspects of the music. An alternative approach which has worked very well for many instruments is to incorporate additional symbols into the orthochronic system without altering its basic characteristics. Thus, for instance, many piano scores contain small numbers over some of the notes, which prescribe fingering. There are also many simple symbols which can be placed over or under notes to specify such things as the nature of attack, loudness, phrasing. The spatial constraints here are rather indefinite. If there are too many of these additional symbols the score takes on a cluttered appearance, and there is a danger of overloading the reader with detail. Compare Figures 6 and 7 which are different editions of the same music. Figure 6 is crowded with specific performance directions. Figure 7 uses such directions sparingly. In general, editors must take a middle line between giving the performer no help at all (trusting to his own musical competence) and giving him so much that some of it is bound to be ignored. The trend recently has been towards the former approach, at least in respect to classical music. It is not coincidental that Figure 6 predates Figure 7 (a contemporary edition) by some decades.

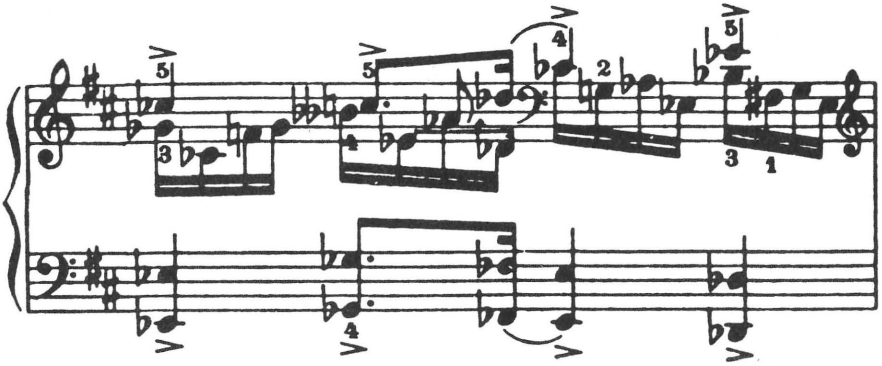

Figure 4 A bar from F. Chopin, Sonata Op. 58, in orthochronic notation. Source: Karkoschka (1972). p. 12.

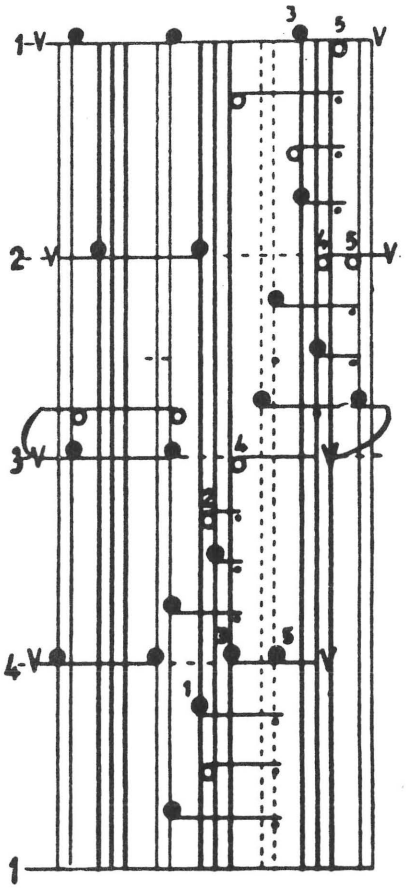

Figure 5 The same music as Figure 4 notated in Klavarscribo. Source: Karkoschka (1972). p. 12.

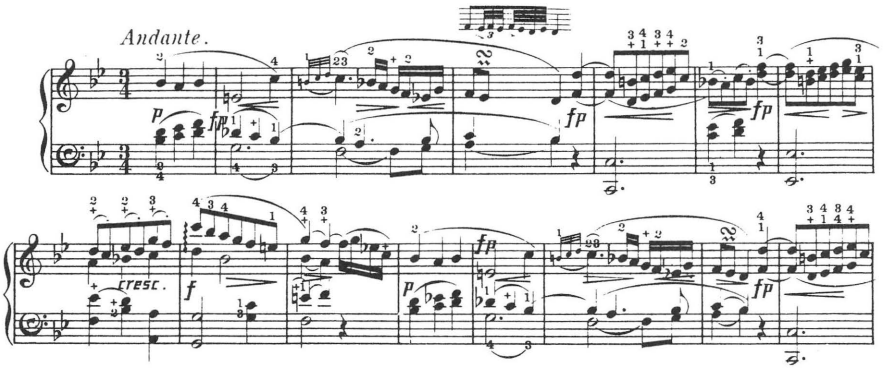

Figure 6 Opening bars of "Andante" from W. A. Mozart, Sonata in F, K. 533. (Ed. A. Zimmerman). London: Novello. Undated (c. 1900).

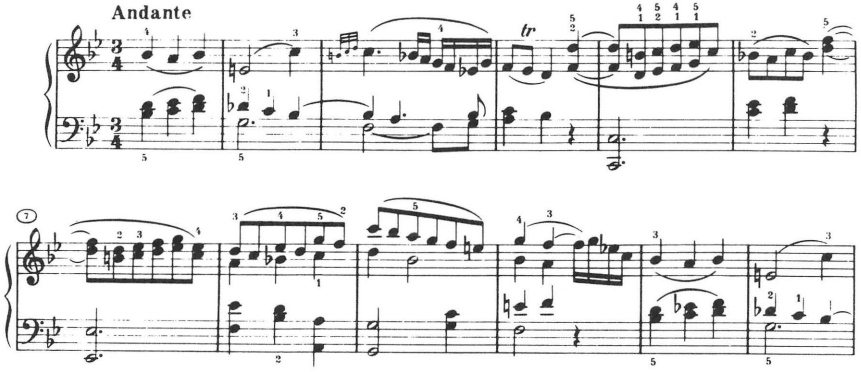

Figure 7 The same music as Figure 6. (Ed. W. Lampe). Munchen: Henle, 1955. Reproduced with the publisher's permission.

3. The staff

The most noticeable and characteristic feature of any segment of orthochronic notation is the grid of five horizontal lines on which the notes are placed. These lines are called staff-lines, with the five lines known, collectively, as a staff (plural staves). Unlike the tablatures, the function of the staff in orthochronic notation is to represent the pitch dimension. From earliest times pitch was seen as one of the most important aspects of music to notate. The earliest notations were alphabetic, dating at least as far back as pre-Christian Greece. In these systems each note of the scale was designated by a letter or other symbol. To notate a melody one simply wrote out the letters in the correct order. The first use of space to indicate pitch was the development of *neumes*

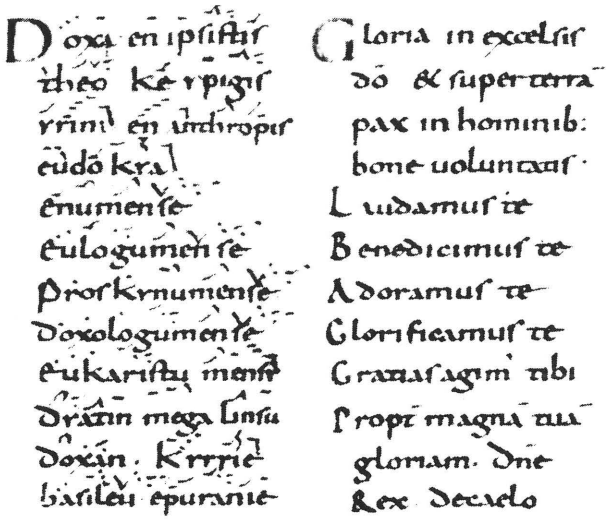

Figure 8 "Gloria" from a St. Amand manuscript c. 871. Paris: Bibliotheque Nationale, ms. lat. 2291. Reproduced with the Trustees' permission. Source: Abraham (1979), p. 62.

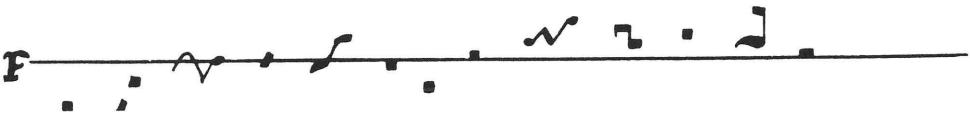

Figure 9 Tenth century neume notation (author unknown). Source: Read (1974), p. 10.

in around the eighth century AD. These were used almost exclusively in the liturgy of the Christian Church to notate the plainchant melodies to which prayers were sung in religious communities. Figure 8 shows one of the earliest examples of neumes from a French manuscript C 871. The neumes are written above the Greek text and indicate the directions of pitch change. Thus ⟋ denoted a rise, ⟍ a fall, and ⟋⟍ a fall followed by a rise.

This system had no power to specify the extent of a pitch change, only its direction. Read (1974) comments "Useful solely as a reminder to the singers, refreshing their memory of a general rise and fall originally learned by rote, the neumes provided only the over-all contour of the plainsong melodic line – not an exact map by which a novice might approach an unknown musical territory" (p.7). This comment introduces a very important consideration, that of the context in which a notational system is used. Neumes did not do badly what orthochronic notation does better. They probably served the cultural and psychological requirements of their users better than modern notation would have done. They had an economy of symbolism which exactly served its purpose within monastic choirs, where novices would learn most of what they

needed through repeated hearing of the same hymns. It was only, possibly, as chants elaborated and diversified that more precise mnemonic systems were necessary.

The next conceptual development was the use of vertical distance to indicate the extent of the pitch change. Figure 9 shows an early example from the tenth century which places neumes relative to a single horizontal line given a precise pitch (*F* below middle *C* in this example). The distance of a neume from this line gives an indication of its pitch distance from *F*. The difficulty with this, from a modern musician's point of view, is that the y-axis is not calibrated. A reader cannot easily tell what range of pitches is covered, nor the exact distance of any particular neume from the baseline. Both these problems were solved over the next two centuries by the addition of further horizontal lines denoting other precise pitches. By the 12th century the number of lines had reached 4. This number seems to have been supported by two principles. One was that lines were provided for alternate pitches in a scale. Thus, the bottom line might represent *A*, the space above it *B*, the next line *C*, the next space *D*, and on. This principle remains in operation today in orthochronic notation. It means that the pitch of a note need never be estimated by reference to its physical distance from another point on the paper. It can be identified precisely by counting up the staff lines. For experienced readers, even this is not necessary. The position of the note is directly known by a kind of spatial subitising. The look of each position on a staff becomes so distinctive that no conscious counting is required.

The second principle was that the number of staff lines was tailored to the pitch span of the melodies. Most melodies could be encompassed within nine successive pitches of a scale. This span (about an octave) would be the comfortable range of an average untrained male voice. Even in the 12th century most notated music was church music sung exclusively by monks. As time passed, however, it became more customary for instrumental music to be noted on a staff. Instruments have larger pitch spans than voices, and so more lines were needed to encompass the melodies that could be played on instruments. Six lines were fairly common and up to fifteen were known. Somehow things settled down and the five lines of modern orthochronic notation became the norm. Although it is impossible to know this with certainty, it seems probable that one of the reasons why the five-line staff survived is that it represents the best compromise between good span and readability. Figure 10 presents a ludicrously extreme example of the readability problem. This is part of a composition by E.Brown (1952). There are fifty staff lines, and it is practically impossible for a reader to keep any bearings at all. Luckily for the performer, Brown does not intend the lines to be taken seriously, but the moral is clear enough. There is a limit to the number of lines a staff can contain if a performer is to use it to rapidly identify the pitches of successive notes. This presents a problem within orthochronic notation when melodies span more than eleven notes of a scale, as frequently happens. The problem has been solved in three ways, all in common use today.

The first solution is to recalibrate the staff as the melody rises above or falls below the range of the staff. This is conventionally done by arbitrary symbols known as clef signs. The leftmost symbol in Figure 1 is the treble clef sign, and it indicates that the

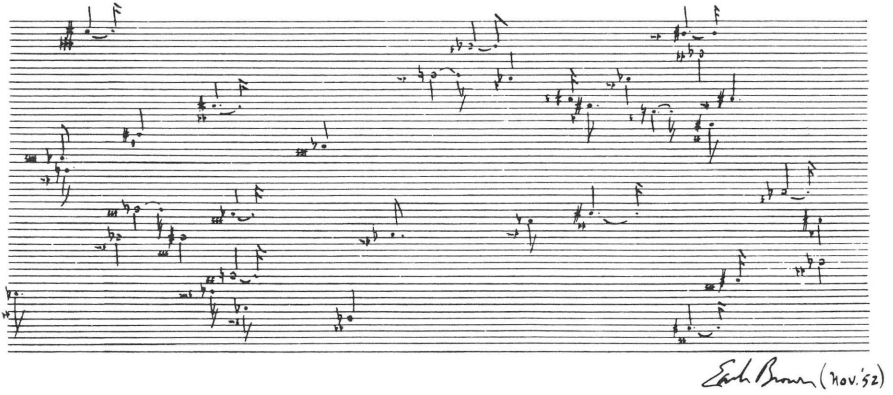

Figure 10 "November 1952" from *Folio* by E. Brown. New York: Associated Music Publishers Inc. Reproduced with the publisher's permission. Source: Karkoschka (1972) p. 90.

bottom line of the staff is to be read as the *E* above middle *C*. There are four or five such symbols in common use, and when one of them occurs in the course of a piece of music it instructs the reader to recalibrate everything from that point onwards until another clef sign is encountered.

For tunes which skip about a lot, clef signs are, however, rather unsatisfactory. If the clef must be changed every few notes the reader will spend as much time interpreting the clef symbols as the notes themselves. Another solution which is in common use is to temporarily extend the staff by ledger lines. The first note of Figure 1 is on a ledger line and the reader is asked to interpret it as if it were another staff line below the lowest one shown. In fact, the reader may imagine the staff is extending indefinitely both above and below its visible portion, but to avoid the problems associated with Figure 10 only that portion of it becomes visible which allows a single note to be identified. Thus, ledger lines extend only as far on either side of a note as to make them clearly visible. Most fluent readers become adept at rapidly interpreting up to five ledger lines. With more lines, problems of readability again develop. Figure 11 shows two extracts using ledger lines; (a) would be easily read by most competent readers today; (b) would not be fluently read.

At this juncture a comment about the vertical spacing of staff lines and ledger lines is pertinent. Although not logically necessary, it has always been standard practise to make the distances between adjacent staff-lines equal to one another. This has meant that the vertical distance between notes is an accurate measure of the pitch distance between them.

Although this information is technically redundant, it does provide readers with an additional analog cue to pitch which arguably supplements the ''digital'' information supplied by the lines themselves. There is, for instance, some evidence that music

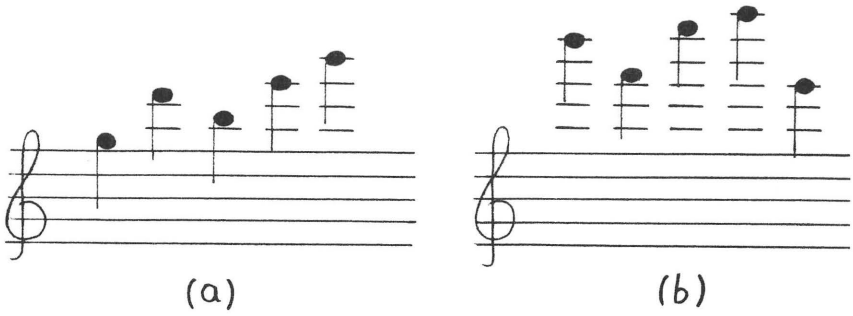

Figure 11 Two examples of the use of ledger lines: (b) is more difficult to read than (a).

readers can be aware of the contour of a notated melody before they can precisely identify the constituent notes (Sloboda, 1978b). It is unclear how much use competent musicians make of this information in performing situations, but one informal observation supports its importance. It sometimes happens that staves are printed so close together that there is a difficulty about placing a note on, say, the fifth ledger line. If the spacing of the ledger lines were to preserve the spacing of the staff lines, they would come too close to, or overlap, the adjacent staff. In such a situation the ledger lines are squashed closer together. This, of course, upsets the analog representation of pitch, and many musicians complain bitterly that such notes are very difficult to read, especially if the music is unfamiliar. In the best publishing houses this crowding of ledger lines never occurs. The solution is, of course, to move the staves further apart, but with fewer staves to the page, production expenses increase. This probably accounts for the persistence of this troublesome practice. An allied point concerns the size of a staff. This has both psychological and economic implications. If staff lines are too close together the reader has problems of discrimination, especially if he must be at some distance from the score when performing. Conversely, if the lines are too far apart then it may be impossible to see the stave clearly with a single fixation. Similarly, from the printer's point of view, closely spaced lines demand finer (and thus more expensive) printing. Widely spaced lines mean fewer staves per page and thus more pages. Engraving tradition seems to have settled on a between-staff-line distance of about 0.20 to 0.25 cm as suitable for most performing scores.

The third solution to the problem of increased melodic span we owe to the rise of polyphonic music in the middle ages. Polyphonic music is that where several streams of notes are performed simultaneously. This type of music can be vocal, where different singers take different melodies simultaneously. It can be for an instrumental

group, where, for instance, different pipe instruments play different notes. Or it can be for a keyboard instrument played by one performer producing simultaneous sounds with different hands and fingers.

In all these cases, the notational problem is to indicate which person or hand takes which stream of notes. To put all the notes on to a single staff would be to overcrowd it hopelessly if more than a couple of parts were involved. So the solution arrived at fairly early was to give each part a separate staff. These staves were vertically aligned and joined in some way to indicate simultaneity. A typical example of this is given in Figure 4. This is part of a piano piece – the curved bracket at the left indicates that the two staves are to be played simultaneously, the upper by the right hand, the lower by the left hand. This arrangement is necessary when a single performer must read both staves simultaneously: the relevant portions must clearly be as close to one another as possible. It also provides an ideal vehicle for extending the span of pitches which can be notated. This is because the two staves can be calibrated differently by clef signs, as in Figure 4. Suitably calibrated, the piano score has over twice the pitch span of a single staff. In the notation of symphonic music this system can be extended indefinitely, twenty staves being a not uncommonly large number for a conductor's score. Of course, in this case many of the staves overlap closely in pitch calibration, or, indeed, are identical. Each staff is identified with the name of the appropriate instrument at the left of the page.

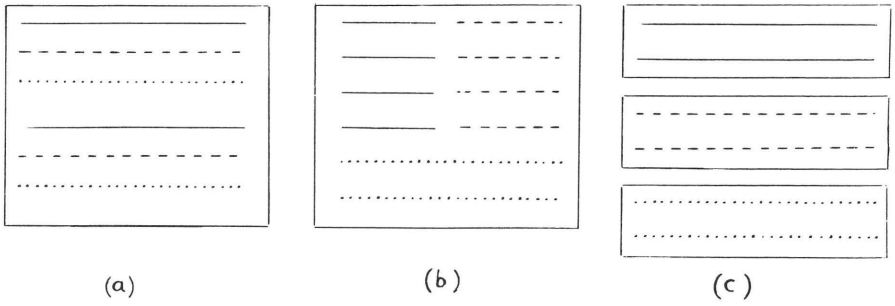

Figure 12 Three arrangements for polyphonic scores.

It remains to add that this score arrangement is not the only possible or desirable one for polyphonic music. When many performers are involved it often makes sense to arrange things so that each performer has primary access to his own part. Figure 12 shows in diagrammatic form the three major ways in which a polyphonic composition can be arranged in space. In this example three parts are shown, but the same principles can apply to any number. The full, dashed, and dotted lines represent the staves for the three separate parts or instruments; (a) is the score arrangement just described, and would be typified by any organ score (separate staves for right hand, left hand and feet); (b) is an arrangement whereby each part is confined to a circumscribed area on the page. Although the other parts would be visible to a

performer they would not get in his way. This arrangement is typical for one piano, three players music and for much medieval vocal music where each part was to be taken by one singer. In (c) each part is physically separate and can be distributed around the room among the various performers. This is the usual format for modern orchestral players where their own part may be one of as many as fifty different parts going on simultaneously. Of these systems, (a) is undoubtedly the oldest. It allows the reader to keep a direct check on the synchrony of the parts. Examples (b) and (c) make visual checks progressively more difficult. The survival of (c) depends upon the development of cues to synchrony. One, outside the scope of this article, is the conductor, who provides an external reference for a group of performers through his gestures. Probably more crucial was the development of methods for notating cues to synchrony and timing within individual parts. It is to this second major dimension that we now turn.

4. Bars, timing and rhythm

We have seen how orthochronic notation has commandeered the vertical dimension for notating pitch. It is only a first approximation to say that the horizontal dimension represents time. Temporal aspects are certainly represented, but not in any strict analog sense. Distance from left to right is not a direct measure of time elapsed. To understand how time notation works it is necessary to return to its beginnings in history.

As we have seen, most of the early notations were designed for singing liturgical chant. Thus there were two parallel notations – one giving the words in an orthography essentially that of today – the other, usually above the words, giving musical directions. In addition, the early music was all homophonic. No matter how many people sang, they all sang the same melody. It would not be unfair to say that up to about the thirteenth century the only timing information explicitly supplied by the musical notation was that of order; the order of the successive words and the pitches to which they were to be sung. If there were differences in the duration of the notes then these were possibly suggested by the flow of the words, or were part of an oral tradition of timing which had grown up in particular communities. Thus, while pitch notation continued to develop towards the modern staff system, the temporal aspect remained undeveloped until the thirteenth century. At that time notational distinctions between long and short notes began to be made.

For some reason, the strategy of using space to indicate time was not adopted. A compelling explanation for this is that scribes felt themselves constrained by the parallel vocal text. From the appearance of Figure 13 one could guess that the scribe had written out the words before embarking on the music. The spacing of the letters and the words is very regular; that of the notes is quite irregular. Clearly, the main consideration was to ensure that the right notes fell over the right syllables. This is what determined the spacing of notes, and so timing had to be indicated in some other way. In general, this was done by altering the shapes or appendages of the notes. In

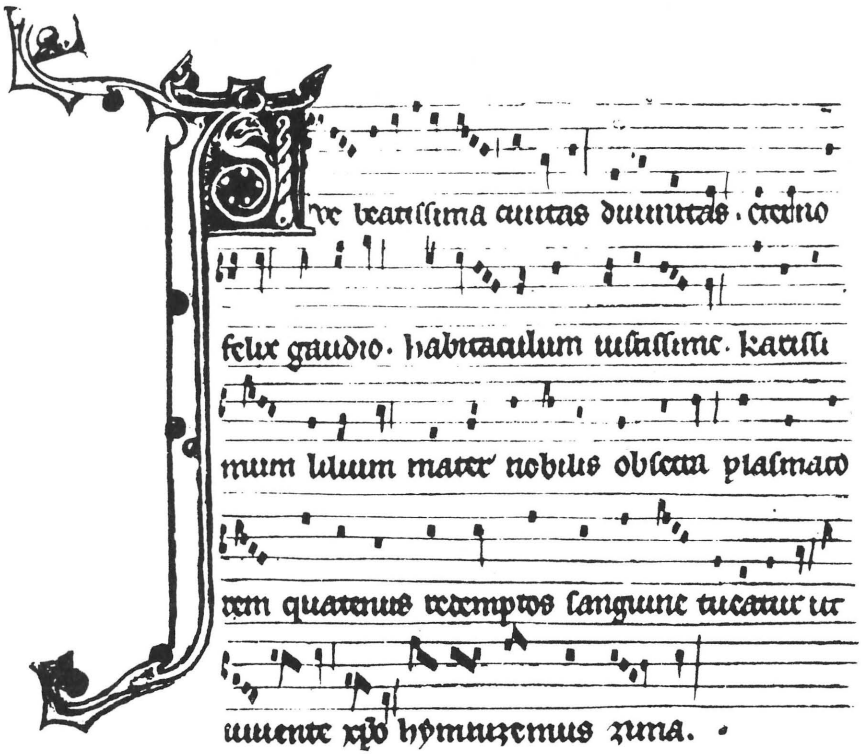

Figure 13 "Ave beatissima" from the *Codex Montpellier* (thirteenth century). Montpellier: Bibliotheque de l'Ecole de Medicine. Ms. H 196 p. 93'. Reproduced with the Trustees' permission. Source: Apel (1953) p. 291.

this example, from a thirteenth century French manuscript there are three types of note: squares with stems, squares, and diamonds, denoting long, short and shorter notes respectively. It is along this path that rhythmic notation continued to develop despite the fact that much notated music now came to be instrumental, and so without constraint of words.

The history of the subsequent development of rhythmic notation is of some complexity, but certain principles emerged which have remained with us to the present day. One was that note durations were conceived as simple multiples of each other. The other was that simple combinations of durations would often recur several times thus setting up a rhythmic grouping. A simple example of this, as apposite to the twentieth as to the thirteenth century, is given in orthochronic time notation in Figure 14 (a). In this example the white (unfilled) notes have twice the duration of the black (filled) notes. If the duration of the black note is taken as one unit, then we can see a repeating pattern every three units. This is expressed today by saying that a triple

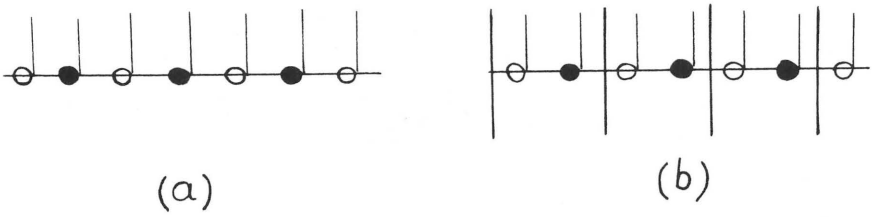

Figure 14 Rhythmic notation.

metre is being used. When the rhythm is more elaborate, it is not so easy to determine what the repeating grouping should be just by examination of the note values. Sometimes a sequence can be grouped in more than one way. In these cases, some notational means of specifying the grouping *directly* is required. It is required because rhythmic groupings have important performing implications. For instance, the first note in each rhythmic group is typically accented in some way (e.g. played louder). This is particularly important for communicating the rhythmic structure to a listener. A direct and visually compelling cue to rhythm is particularly useful when the musician is reading unfamiliar music "at sight". Under these conditions he will not have time to work out the rhythmic structure from the note sequence itself. In more leisured study circumstances, or where the music is already partly familiar, direct cues are not so crucial. In modern times sight-reading is required of musicians in many circumstances. The same was not true 700 years ago. The possibility of effective sight-reading has rested upon the emergence of two notational devices.

The first of these is an indication, at the beginning of a sequence, of the number and type of notes in each metrical grouping. This is achieved by two numbers (see Figure 1). The upper number tells the reader whether the recurring grouping is 2,3,4 or whatever number of units. The bottom number specifies the unit. These two numbers together comprise the time signature of a piece, and in Figure 1 this signature indicates a triple metre, with the filled note (known as a quarter-note in America) as the unit.

The time signature, however, leaves a very important question unanswered, for it does not tell the reader which note is the first, accented, note in each group. One cannot simply assume that the first note printed is given the accent. Many tunes begin on an unaccented note. Furthermore, it would be quite difficult for a reader to keep track of the metre through a long piece without dropping a note. A recurrent cue is required. This is supplied, in orthochronic notation, by bar lines. These are vertical lines, running the extent of a staff, which precede the first note in each metrical unit. A bar is the space between two adjacent bar lines, and so each bar contains a complete metrical unit with its first, accented, note at its leftmost edge. The correct barring for Figure 14 (a) is shown in Figure 14 (b).

Regular metrical barring did not become widespread until the sixteenth century, but when it did if offered a number of cues to synchronisation. The first is the use of bars

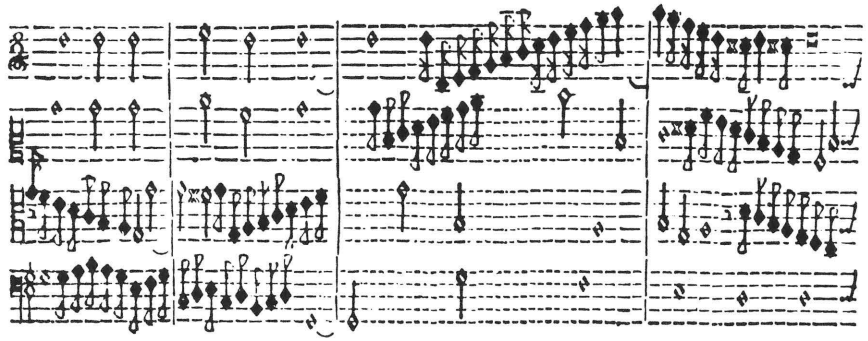

Figure 15 From A. Mayone *Primo libro di diversi capricci per* Naples, 1603, p. 70. Source: Apel (1953) p. 17.

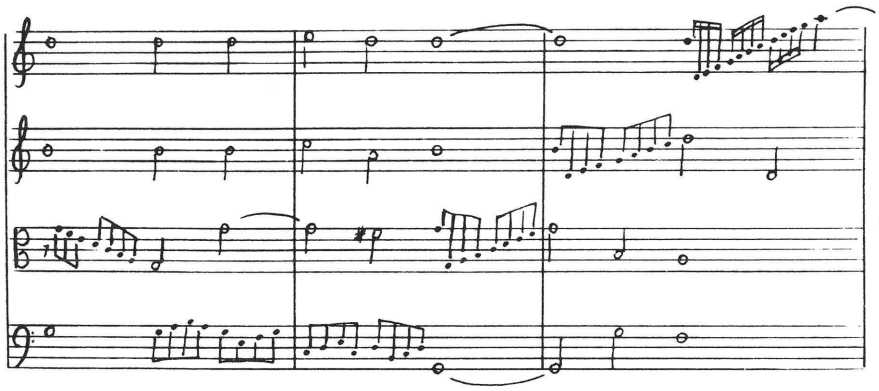

Figure 16 Transcription of first three bars of Figure 15 using the conventions of modern orthochronic notation.

as a tally in part performance (Figure 12 (c)). If a performer knows the metre and the speed he can keep track of the music by counting the bars. In symphony orchestras a primary responsibility of a conductor is to maintain the beat, moving his arms in such a way as to clearly indicate the timing of the first note in each bar (the down-beat) and the regular progression of metrical units during each bar. Performers may then tally each downbeat with a bar on their score. In addition, many parts actually number the bars so that in rehearsal performers can agree to 'start from bar 50'.

The second cue to synchronisation is available only in parallel score arrangements (Figure 4 and 12(a)). This is achieved by vertically aligning bar lines in the separate parts. In many instances the bar line will be ruled right down the whole system of parallel staves to make more salient the points of synchrony. Figure 15 shows an early example from Italy (Naples, 1603). Two features are of note. One is that when some

parts contain many short notes the bars must be widened to accommodate them legibly. Thus bars are not of regular size. The trend in modern orthochronic notation has been to attempt to minimise difference in size between successive bars, although it is not possible to eliminate them entirely. In some music, a bar may contain so many notes that it requires the whole width of a page. To maintain this width for bars containing only one or two notes has little psychological advantage, and severe economic disadvantage. The other feature is that the distance between adjacent notes in parallel parts cannot be kept constant if the bars are to be aligned. The clearest example of this in Figure 15 comes in the third bar, where the top two parts have many closely spaced notes, whilst the bottom two parts contain only few notes. In this example, and in general up to at least the eighteenth century, no consistent method for spacing out the notes within a bar was adopted. Thus, although bar-lines provided conceptual points of synchrony for a reader, one could not drop a plumb-line at any other point within a bar and expect any synchrony between the notes encountered in the various parts. The time dimension from left to right was stretched and contracted quite arbitrarily within each bar. It is worth noting that in some early scores the bar lines were not even straight. What appears to have happened is that bar-lines became so useful to musicians that scribes often went back over old un-barred scores to put the bar-lines in. In these older scores there was often no attempt whatsoever to maintain vertical alignment of parts, and scribes were apparently too hard-pressed to write out the music afresh. So one finds examples like the splendid British keyboard score (circa 1540) in Figure 17, where the clumsy barring indicates the extent of the visual splits which early keyboard players were required to perform. Notational clarity apart, a modern keyboard performer would find it impossible to sight-read such a score. I suspect that a sixteenth century player would have found it equally difficult, and it suggests that such a score would have had rather different purposes—perhaps archival, perhaps as an aid to a performer attempting to memorise a piece, or perhaps as a starting point for an improvisation.

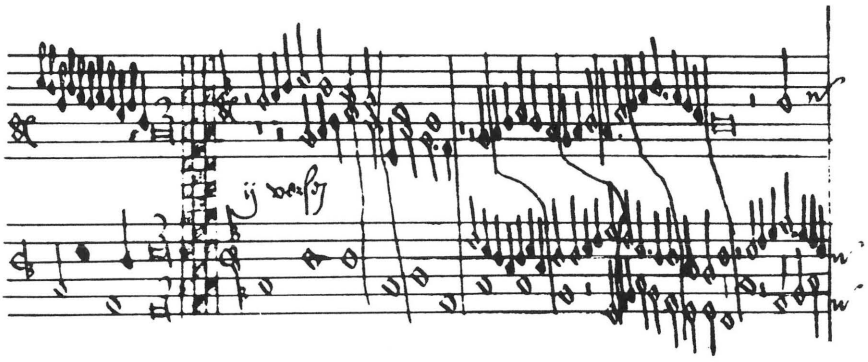

Figure 17 From a keyboard manuscript. London, 1540. British Museum, Ms. Add. 29996, p. 117'. Reproduced with the Trustees' permission. Source: Apel (1953) p. 11.

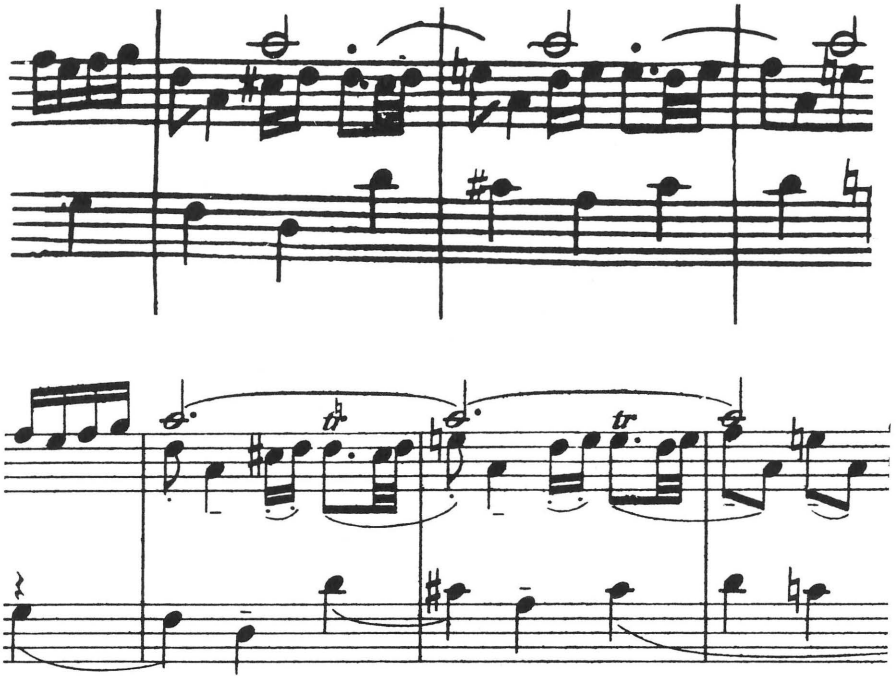

Figure 18

(a) From G. F. Handel, Organ concerto Op. 4 No. 3. Publisher unknown, c. 1800. Source: Cole (1974) p. 58. (b) The same music as (a). Ed. K. Matthaei. Kassel: Barenreiter, 1956, p. 59. Reproduced with the publisher's permission.

It was not really until the nineteenth century that what we now see as a logical corollary of bar-lines, namely proportional spacing of notes, came to be widely accepted. It is now standard notational practice to make systematic use of the space within the bar. Two main principles summarise this practice:

Firstly, notes which begin simultaneously must be vertically aligned. This has the useful consequence for, say, a pianist, that all the notes he must play simultaneously can be seen from a single vertical slice of score. This supercedes an earlier, intermediate, principle of spacing whereby a note occupied the centre of the time-space allotted to it, rather than the left-hand boundary of that space. This meant that a keyboard player had to scan horizontally as well as vertically to find the notes which had to be played together. Figure 18 shows a small extract from a score of a Handel organ concerto; (a) is an early printed score (c. 1800 ?); (b) is a modern reprinting. The clearest example of the change can be seen at the beginning of the

second bar. In (b) the three notes to be played simultaneously are precisely aligned; in (a) there is considerable divergence, particularly for the top note, which, because it lasts for the whole bar, is placed almost in the middle of the bar.

Secondly, the space between a note and the following note must be proportional to the time between their respective onsets. This has always been interpreted rather loosely. It is not necessary that the space following a particular note need be twice that of a note with half its duration, only that the space be discernibly larger to the reader. This principle, in fact, follows as a necessary consequence of the first principle applied to parallel staves. It is, however, used nowadays, even in cases where the constraints of alignment do not strictly require it. Thus, as in the pitch dimension, analog spatial information about timing is available to the reader as a formally redundant cue. Again, what psychological evidence exists is consistent with the idea that, at least in some circumstances, readers make use of proportional spacing of notes in organising their perception of the score (Sloboda, 1977). To illustrate the principles of proportional spacing in an extended orthochronic example, I offer as Figure 16 my own transcription of the example in Figure 15 (first three bars).

5. How well does music notation use space?

The preceding sections outlined the principal spatial characteristics of orthochronic notation and its forbears. In this final section I wish to turn to questions of greater generality: to look at music notation not from the point of view of musical history, but from that of the formal study of symbolic systems. This shift in perspective allows us to focus less on questions of development and more on questions of information content, symbolic structure and psychological effectiveness. Any musical notation system must code certain types of information about the music. It uses particular symbols in particular spatial arrangements to represent this information, and has a greater or lesser effectiveness in conveying this information to the reader in a manner consistent with his requirements. These issues can be examined without reference to the historical context of particular notational systems. What follows is an outline of the way in which these issues may be expanded and articulated to form a conceptual framework within which systematic scientific evaluation of alternative notations could be carried out. It is necessary, I believe, to say something about these three issues separately, although the overall goodness of a system will be the result of a combination of factors from the three levels.

(a) *Information content*. The information conveyed by different musical systems can vary in two major respects: how specific the information is, and what type of information is conveyed. The notion of specificity is best conveyed by two extreme examples. Figure 19 is the score of a contemporary composition in which the performer is asked to contemplate what he sees and play anything that the pattern suggests to him. No detailed rules are supplied for interpreting the various elements. Thus there is no sense in which particular elements of the display correspond to particular sounds or features of the music. At the other extreme are notations which

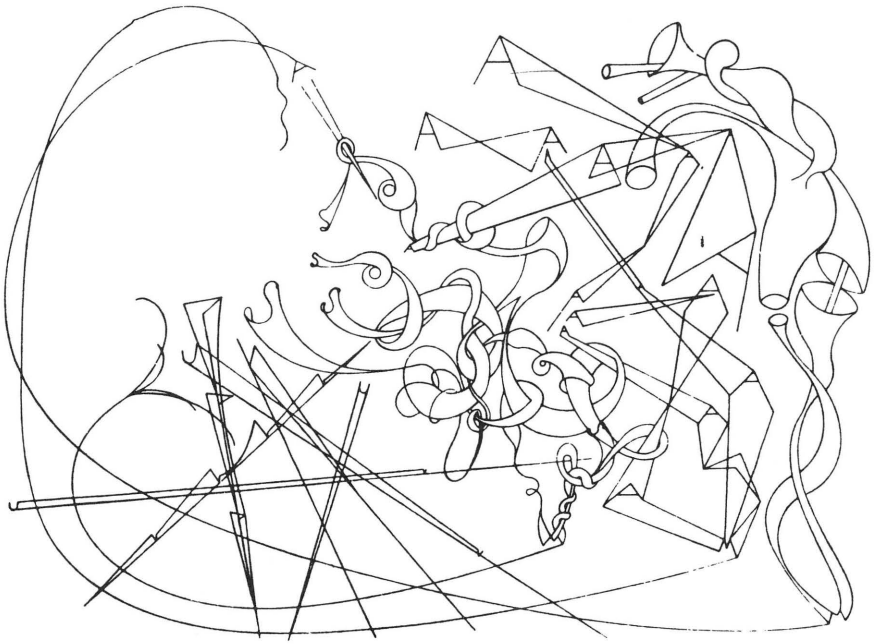

Figure 19 Sheet 3 of A. Logothetis, *Cycloide*. Munich; Edition Modern, MusikVerlag Hans Werwerka, 19. Reproduced with the publisher's permission. Source: Karkoschka (1972) p. 129.

almost totally specify each sound. Some modern electronic compositions achieve such specificity by giving the precise details of all the oscilators, filters, etc. Orthochronic notation is specific in that it provides information about each note in a piece of music. Of course, even in such a case, specificity can vary according to how *much* information about each note is provided. Sometimes only one or two dimensions of the sound (e.g., pitch and duration) are notated. Sometimes many more dimensions of the sound are specified. The need for specificity poses certain formal problems for a notation system. One is how to represent the order of events. Another is how to correlate the elements in the various dimensions which refer to the same note. How do we show, for instance, that a given note has a pitch x, a duration y, and an intensity z?

At a given level of specificity there can be further variation in the *type* of information represented. One distinction of type was mentioned in Section 2, that between systems which notate actions (e.g. the tablatures) and those which notate aspects of the sound. If any consideration were to limit the number of dimensions which could be notated, then the different types of information would have to be traded of against one another. Increased specificity in one area would imply decreased specificity in another.

(b) *Symbolic structure*. By structure is meant both the nature of the symbols chosen and their spatial arrangement. Design of the structure becomes a formal problem when some degree of specificity is required.

There are a whole set of solutions to this problem which we could characterise as essentially non-spatial. As an example, we could assign each event a number (1 being the first event, 2 being the second, and so on) and then make certain statements about the various characteristics of each event (e.g. '2 has a pitch of 565 Hz', '7 has an intensity of 65 dB', etc.) These statements could be arranged on the page in an arbitrary order. One's immediate feeling that such a system would be pointless (i.e. ineffective) does not invalidate it as a perfectly consistent and correct way of symbolising the specified dimensions of the music. Nonetheless, a notation system could utilise the space provided by the writing surface in a more systematic fashion. The most primitive use of space is to order the elements in some way. With music, the ordering is nearly always temporal. Information about the first note is spatially adjacent to information about the second note, and so on. This is, presumably, because temporal order is the dimension which *makes* music. In the example just given, one could achieve this characteristic by grouping all the statements about event 1 together at the top of a page, with all the statements about event 2 following immediately below, and so on. This could be structurally described as a list. The next advance on a list is a matrix. Here, the two spatial dimensions of a page are utilised systematically. Thus, for instance, the horizontal dimension might represent the temporal order of events, with the various characteristics of each event separated out vertically down the page. So, for example, reading across one row would give the pitches of successive notes, reading across another row would give their intensities, and reading down a column would give all the features of a single event. (I owe much of my conceptualisation of structure and some of the terminology to Twyman (1979)).

A different use of the two dimensions might be called 'co-ordinate'. Here, both horizontal and vertical dimensions represent ordinal axes on which some 2 dimensional function may be plotted. Orthochronic notation can be seen, in part, as a graph of the pitch (y)–time (x) function, each note plotted as a point on the graph. Matrix and co-ordinate notations can be made very powerful by the use of two other supplementary devices, partition and clustering. In partitioning, one divides the page up into discrete areas where space may serve a different function. Thus, for instance, in Figure 15 the pitch dimension is not represented continuously in the vertical dimension. It only operates within each staff, recalibrated for each. The space between the staves has no pitch implications at all, and so other, pitchless information may be placed here. In some modern scores, for instance, a rising and falling line between staves is given to indicate increases and decreases in speed. A most imaginative use of partitioning is shown in Figure 20. This is the second page of Busotti's *Siciliano*. Here, vertical space within staves is interpreted conventionally, but the slant of a staff indicates the degree of acceleration or deceleration of speed. Dotted vertical lines indicate points of synchrony.

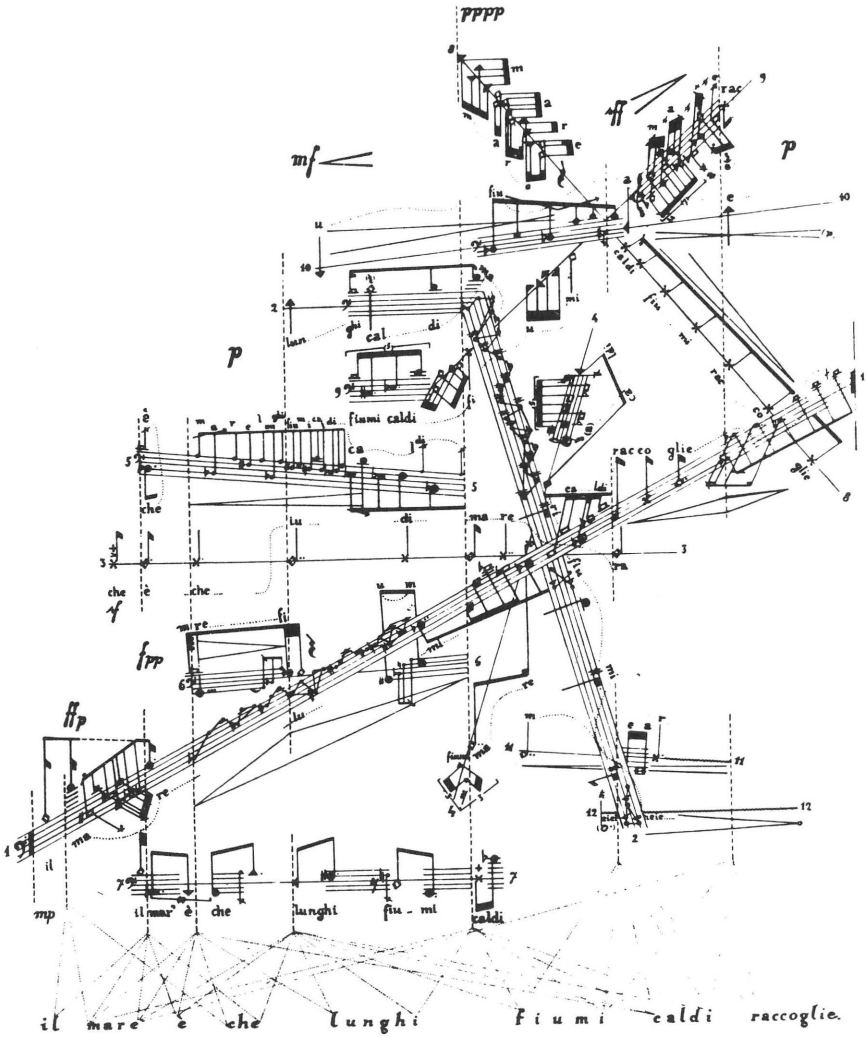

Figure 20 Page 2 of S. Bussotti, *Siciliano*. Florence: Aldo Bruzzichelli, 19 . Reproduced with the publisher's permission. Source: Karkoschka (1972) p. 94.

Clustering refers to the practice of representing several dimensions of an event, not by separate symbols, but by different aspects of a composite symbol. Thus, a note's position could indicate its pitch, whilst its shape indicated its duration, and its colour indicated its intensity. This is a useful economy of space, which allows one to compact several rows or columns of a matrix into one.

These strategies, in their detailed operation, would seem to account for most of what one could say about the possible structures of music notation systems. When a combination of all the strategies is available there would seem to be very few formal limits on the number of specific dimensions which could be represented by a symbolic system. If questions of how well space is used have significance at this level, it is with respect to the economy of space. A system which allows clear representation of twenty pieces of information has a claim to greater efficiency than one which allows representation of only ten pieces of information within the same area of space. But for a criterion of what constitutes a clear representation, we must turn to the final, psychological level of analysis.

(c) *Psychological effectiveness.* Music notations are for use by readers. Their effectiveness is the extent to which readers are able to retrieve the information about the music which they supply. Effectiveness is not simply a function of the symbolic structure of the score, nor of the information it contains, but of the conditions under which a reader is expected to retrieve the information. These conditions may include general factors of human psychological functioning; they certainly include factors specific to the particular reading situation. All I am able to do here is to enumerate a representative sample of issues pertaining to psychological effectiveness. Although these issues are all empirically determinable, almost none have received rigorous empirical treatment.

A fundamental issue I have already referred to is the extent to which a reader is expected to provide a coherent performance of the music at first sight. Clearly, the more the reader must do without prior knowledge, the more constraints are placed upon what is an acceptable notation. Some of these constraints are straightforward matters of legibility. Even within a single notational system clarity and consistency can vary greatly according to the care with which a score is laid out and the nature of the reproduction process. It is distressing to discover that many of the parts which players in major symphony orchestras must use (at least in Britain) are shockingly presented and reproduced. Figure 21 shows a not untypical example of the kind of thing an orchestral player must put up with. It has not been determined how much this upsets performance. Certainly the players themselves complain bitterly. A major difficulty is spatial uncertainty. This can occur when staff lines are badly printed, or too close together. In this case it is sometimes difficult to see which line a note is centred on. A similar problem occurs when noteheads are too large with respect to the staff lines. A different type of spatial uncertainty is caused when ancillary marks are not properly aligned with the notes, or when they are too far from the staff to allow easy estimation of their alignment.

These considerations lead to the wider issue of discriminability. When a notation system uses a large number of different symbols, two different symbols can often look very much alike. If a reader is expected to discriminate the symbols at speed, then he will arguably be helped by symbols which are as different as possible in appearance. On the other hand, if he is to perceive relations between different notes then he will also arguably be helped by symbols which preserve that relationship (e.g., symbols for notes close in pitch will be more like one another than symbols for notes not

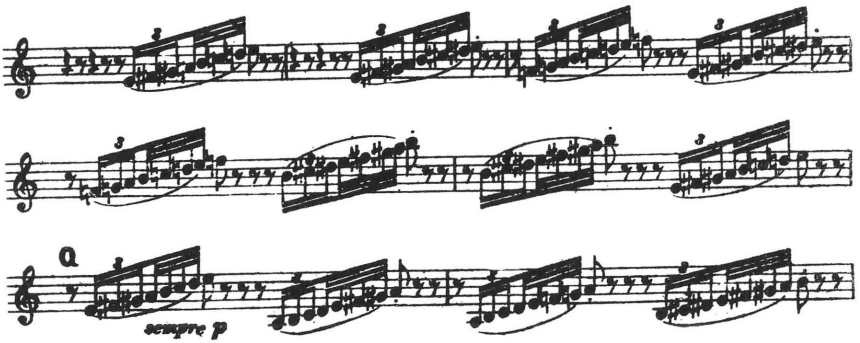

Figure 21 Extract from an orchestral part used by a British symphony orchestra, date and publisher unknown.

similar in pitch). The analog representations of pitch and time in orthochronic notation are based directly on this latter principle. Here, closeness in space maps directly on to closeness in pitch and time. When space is the notational dimension then it is possible to achieve increased discriminability simply by expanding the axes (i.e., make the staves larger and the notes farther apart). But to do this raises problems in a different area, that of the visual span and eye-movements.

If a reader is to use the score to organise immediate performance then he must take in all the information required for a particular action with as few fixations as possible. An average rate of performance for a moderately demanding piano piece might be ten notes per second (some of them executed simultaneously). It would be difficult to achieve more than four or five separate fixations in this time, and so the information required must necessarily be contained in a small area. Expanding the dimensions of the symbols would increase the number of fixations required to read the same number of symbols. Thus an effective notational system also has to be compact, and this poses problems when more than one dimension of the same event must be notated. The conventional matrix would be unhelpful because different aspects of one note could be spread right down a page. The device which arguably saves orthochronic notation is the clustering described in the previous sub-section. By using different visual aspects of the same symbol (position, shape) to indicate different sound aspects of the same note (pitch, duration) one can increase the visual density of the information at little expense to clarity. In general, a system such as the staff system would seem to be able to support about five different dimensions of sound without difficulty. For instance, the two spatial dimensions could represent pitch (vertical) and order (horizontal). Shape of note is a third dimension which could be used to represent duration. Ancillary symbols above and below the note could represent intensity and phrasing. There is, however, clearly some limit to the number of dimensions which could be usefully added in this way. This shows itself in lute notation (Figure 2) where note shape is used to designate which fret to touch, at the expense of a means of signifying duration. Klavarscribo fits in an extra dimension only because the arrangement of

keys on a piano is isomorphic with the pitch dimension. Thus, the same horizontal information specifies both pitch *and* which key should be pressed (Figure 5).

A final requirement of a system that is to be read at sight is universality. One can only become proficient at reading any system if one encounters a large body of music written in that notation. Many contemporary notation systems are ruled out on this count because they are designed by the composer with a particular composition in mind, and are used only for this composition. Thus, Brown's fifty-line staff (Figure 10) has never been used before or since his one composition. Many contemporary notations are designed with the expectation that a performer will devote considerable time to a study of the particular score and its notation before a performance is possible. It would be inconceivable to arrange a performance of Busotti's *Siciliano* (Figure 20) at sight. Neither he nor his performers would wish for that. Similarly no notational system devised prior to about 1500 would have been read at sight. Musical culture just did not require it.

Today, musicians accustomed to reading orthochronic notation at sight become very sensitive to slight changes in notational practice. One simple spatial example concerns the positioning of note-stems (the vertical lines attached to notes). The modern convention is that descending stems are attached to the left edge of a note-head, ascending stems to the right. Some earlier scores reversed this convention. Informationally and structurally this has absolutely no consequences at all, but it *is* psychologically disruptive. Such scores are more difficult to read at sight, and the subjective impression is of something quite wrong about them.

Even when reading at sight is not required, a score's primary function is to assist a musician in performance. In such cases the score may be used to elicit learned response patterns in the correct sequence. This means that the performer must be able to keep place in the score, even if he does not use all the information in it. When detailed prompting is required, he must be able to find the appropriate information rapidly and effectively. Thus, many of the psychological considerations pertinent to reading at sight remain important in less extreme reading situations. Compactness and discriminability are still necessary. Consistency in positioning of information is also important. A reader must be able to know exactly where to go for particular types of information, even if he does not always need them. For instance, in a conventional orthochronic score, information about intensity is always to be found below the staff.

Two concluding remarks are in order. First, I hope to have shown that spatial factors in music notation are of some complexity and importance, and that there are several frameworks within which empirically determinable questions about efficiency of alternative arrangements may be asked. Second, I need to emphasise that nearly all my remarks about the psychological aspects of music reading are based on personal experience and accounts of other musicians rather than on rigorous investigation. Although it is to be hoped that serious empirical work will be undertaken, I should perhaps finish by firing one warning shot across the bows. Efficiency can only be

meaningfully estimated where a reader is as familiar with the system in question as he is with the accustomed system with which the experimental system is being compared. Failure to find an immediate effect of some notational change does not imply that there will be no effect after the months or years of familiarisation that has been received by the accustomed system. Any evaluation of notational reform is going to be a long-term process, and one which will require a considerable commitment from musicians who are required to learn and operate with new systems. Any short-term experimentation is almost bound to lead to the erroneous conclusion that orthochronic notation as it now stands is the best of all possible systems.

References

Abraham, G. *The Concise Oxford History of Music*. London: Oxford University Press, 1979.

Apel, W. *The Notation of Polyphonic Music 900–1600*. Cambridge, Massachusetts: The Mediaeval Academy of America (fourth Edition), 1953.

Cole, H. *Sounds and Signs: Aspects of Musical Notation*. London: Oxford University Press, 1974.

Hyatt King, A. *Four Hundred Years of Music Printing*. London: Trustees of the British Museum, 1964.

Karkoschka, E. *Notation in New Music*. London: Universal Edition, 1972.

Read, G. *Music Notation*. London: Victor Gollancz, 1974.

Sloboda, J. A. Phrase units as determinants of visual processing in music reading. *British Journal of Psychology*, 1977, *68*, 117–124.

Sloboda, J. A. The psychology of music reading. *Psychology of Music*, 1978a, *6.2*, 3–20.

Sloboda, J. A. Perception of contour in music reading. *Perception*, 1978b, *7*, 323–331.

Sloboda, J. A. Music performance. In D. Deutsch (Ed.) *Psychology of Music*. New York: Academic Press, in press.

Twyman, M. A schema for the study of graphic language. In P. A. Kolers, M. E. Wrolstad, H. Bouma (Eds.) *Processing of Visible Language 1*. New York: Plenum Press, 1979.