DataBases of Musical Information

from the

Center for Computer Assisted Research

in the Humanities

(Third Draft 1/95)

525 Middlefield Road, Suite 120
Menlo Park, CA 94025-3443

Tel. (415) 322-7050; Fax (415) 329-8365
ccarh@Netcom.com

©1992, 1993, 1995 Center for Computer Assisted Research in the Humanities

Table of Contents

I. Introduction: Databases of Musical Information 5
Background 7
Concepts and Processes 7
Experimental and Potential Uses 10
Contents and Sources 12
Rights, Permissions, and Licenses 12
CCARH Staff 15

II. Technical Documentation 17
General Principles of Encoding at CCARH 19
Creation of Logical Records 21
Contextual Data Attributes 24
Organization of Logical Records 25
Illustrations 1-9 27
Sound Information (Stage 1) 41
Sample Score (Instrumental) 43
Input Files for Parts (Instrumental) 44
Technical Description of Sound (Stage 1) File Format 47
Graphic Information (Stage 2) 53
Sample Score (Vocal) 55
Input File for Vocal Part 58
Part Output (Instrumental) 61
Technical Description of Graphic (Stage 2) File Format 63

Introduction

DataBases of Musical Information

Background

The concept of databases of musical information was originated by Walter B. Hewlett in the early 1980's. As an engineer and computer scientist, Dr. Hewlett foresaw the value of creating large corpora of machine-readable materials to complement computer hardware and software. As a musician and musicologist, he perceived a need to enable performers and analysts manipulate musical information. Performers might wish to create editions and arrangements; analysts might wish to pose questions that would be impractical to pursue by manual means. Because of his extensive bilateral exposure to both the business world and academic life, Dr. Hewlett anticipated that commercial vendors of software would be unlikely to meet the needs of serious musicians and music scholars. The tasks involved in representing major musical works in the computer with sufficient precision and consistency were too numerous and complex and the anticipated market too small to encourage commercial investment. He decided to put his own numerous skills to use in pursuit of the goal of creating large bodies of machine-readable materials representing the standard repertories of the sixteenth through the early twentieth centuries.

After preliminary investigations, Dr. Hewlett established the Center for Computer Assisted Research in the Humanities as a non-profit facility in 1984. Since that time he has been active as the Center's volunteer director and principal programmer. The Center's other staff members currently include one musicologist, three musical data specialists, and an administrative assistant.

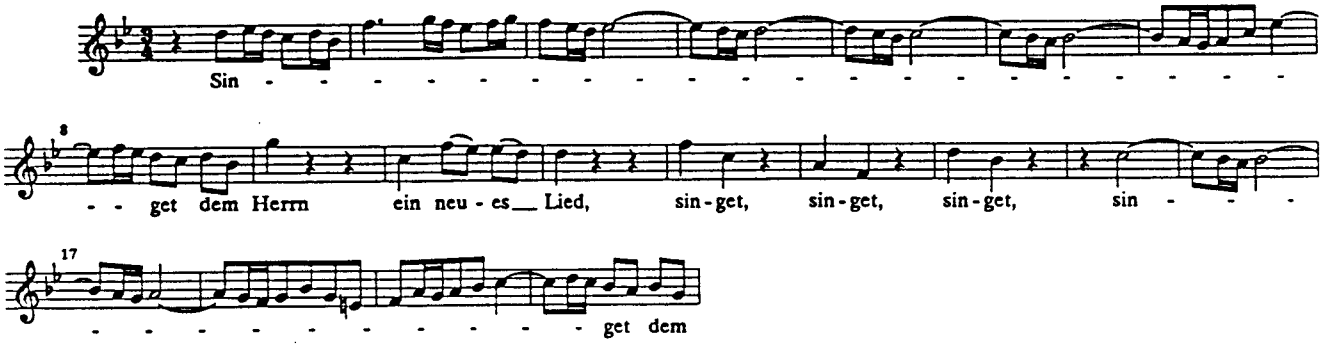
Concepts and Processes

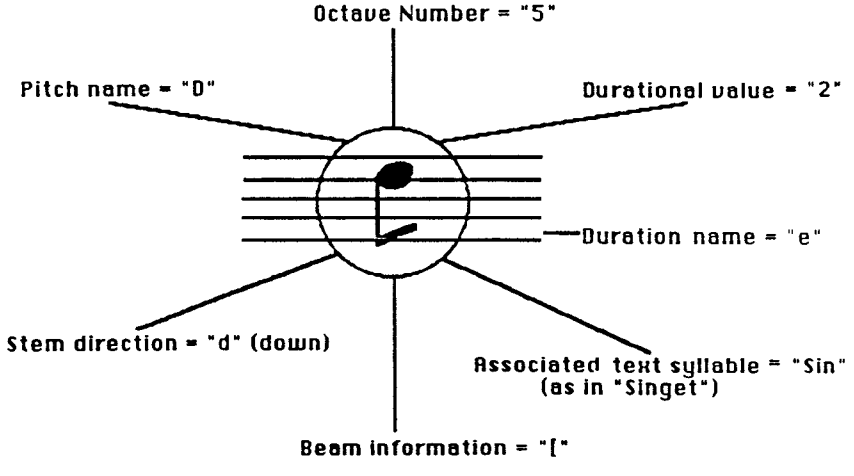
Unlike text, musical notation is not, from a representational perspective, a closed system with a fixed number of characters. There is no standard way of representing musical information or of putting music into a computer, but many arbitrary approaches have proved to be workable. CCARH has played an active role since 1985 in collecting and disseminating information about these in its annual publication *Computing in Musicology*. The Center has thus been in continuing contact with an international community of researchers with related interests and objectives.

The CCARH representation scheme developed by Dr. Hewlett and tentatively called *MuseData*, is designed to facilitate applications of three fundamentally different kinds—sound, graphics, and analysis. The representation scheme is based on alphanumeric codes and is system-independent.

Each note event has a number of attributes. Pitch and duration are the most vital but there are potentially many others. Seven attributes of the first note of the soprano voice of a Bach motet are indicated below:

BWV 225: Soprano





A Note and its Attributes

There are several steps in the entry of a musical work. An electronic keyboard is used to enter pitch and duration data; this part of the process is called Stage 1. All other attributes (text underlay, non-sounding information, general instructions, etc.) are added alphanumerically. This part of the process is called Stage 2. In general, Stage 1 data supports sound and analysis applications, while Stage 2 data supports printing applications. There are certain classes of situations in which the representation for sound/analysis and printing may vary, and in such situations the representation scheme is actually a double one, accommodating both options. Examples of Stage 1 data, Stage 2 data, and the handling of special situations are given in Chapter II.

Since tens of thousands of files are involved in the development of the databases, it is essential that files be consolidated for easy management and compressed for rapid retrieval and searching. Thus it is anticipated that the final state of presentation will constitute a Stage 3. Recent experimentation has demonstrated that the sound and analysis information (Stage 1) for all of the works of Bach could be fitted in a compressed format into 8 megabytes of machine-readable code (that is, on five 1.4 mB diskettes for a standard PC). Further experimentation along this line is required before this format is finalized. It is therefore not documented in this publication.

Each vocal or instrumental part of each movement of each work occupies its own file. Syntax-checking of durational values in the original input is accomplished automatically. After syntax-checking, single files are proof-read before data is stored. Voices are then proof-heard (a musically intuitive method of data verification) in various combinations. Errors and omissions may also be detected when scores are assembled and parts are printed; these activities utilize Stage 2 data.

The pitch and duration (Stage 1) system software used at CCARH gives a significantly more articulate representation of pitch and duration than MIDI (the Musical Instrument Digital Interface) data, which conforms to a standard drafted in 1988. The MIDI representation of music is based on signal processing and was designed for sound-in/sound-out applications to relatively simple music. MIDI's adequacy for handling more complex music and for sound-in/notation-out applications is limited by its lack of support for such things as text underlay, enharmonic notation (C#/Db distinctions), polymeters, and so forth. Stage 1 *MuseData* can be converted to a MIDI representation, but some information may be lost in the process. Extensions to the MIDI standard may lessen possible discrepancies over time. MIDI data can be converted to *MuseData* Stage 1 code, but some supplementary information may be required to bring the data into conformance with other materials. Conversely, MIDI extensions can capture certain kinds

of performance information (velocity, touch, synthetic timbre, etc.) that the CCARH databases do not attempt to capture because the information is not part of the definition of the work. These performance-transcription features would be lost, if present, in the conversion of MIDI files to the *MuseData* file format.

Most systems developed in the Sixties and Seventies for music representation depend entirely on alphanumeric input. Conversion between these schemes, some of which underlie highly successful commercial programs for music printing, and *MuseData* code is generally possible, although few translation programs exist at the present time. Systems for musical notation that are graphics-based (for instance those requiring on-screen assembly of a musical score) are not directly translatable to any other system; in some cases, concealed internal codes, if documented, may be translatable.

Experimental and Potential Uses

CCARH may possess the largest accumulation of fully and consistently encoded, rigorously verified musical data in the world. Our intention is to release it in large quantities by license as soon as we feel confident that (1) no major changes in the representation scheme are desirable and (2) an adequate amount of software exists to search the data.

Most refinements that have accrued to the system of input and representation developed by Dr. Hewlett in the early 1980's have come about in response to needs encountered in use-testing the system. These provisional uses have included the following activities:

(1) Deposit of Stage 1 sample files (the 48 preludes and fugues of J. S. Bach's *Well-Tempered Clavier*) in the Oxford Text Archive in 1986. We have no record of the number of users who have ordered this material, which is available without restriction but should be regarded as provisional. A few minor amendments to the system have occurred since 1986; no documentation is provided by the Text Archive.

(2) Stage 1 files (music by Bach and Beethoven) have been successfully converted to provide a corpus of music for experimentation in electronic conducting, using Max V. Mathews' radio drum (a general description of which is given in the June 1992 issue of *Electronic Musician*), which is under development at Stanford University in collaboration with AT&T.

(3) Stage 2 files (music by Haydn and Mozart) have been successfully converted to the Lisp file structures used in the musical style simulation experiments of David Cope and in diverse analytical implementations under development by Cope and David Jones, both of the University of California at Santa Cruz.

(4) In-house use of Stage 1 files for experiments with computer implementation of simple analytical tasks. Such tasks have included searches for particular sequences of pitches (*e.g.*, B-A-C-H in the music of Bach) and prevalence of particular harmonic intervals (*e.g.*, differentiation of rates of utilization of thirds and sixths in Books 1 and 2 of *The Well-Tempered Clavier*), musical analogues of word-processing capabilities. There has been some exploration of the creation of lexicons of Baroque figural patterns (*e.g.*, in passages consisting entirely of sixteenth-notes in works in duple meter in seven collections of Bach's keyboard music: more than 1500 key-normalized patterns were found). Selected data representing the music of several composers has been used to compare patterns of motion with results reported for other repertories by researchers working in England and Australia.

(5) In-house use of *MuseData* Stage 2 files to create scores and parts for performances of major works by Handel and Telemann. In all cases these have been new editions and in several cases there has been no documented performance since the eighteenth century. The works have included operas, oratorios, and serenatas. The performances have been given by Philharmonia Baroque Orchestra in the San Francisco Bay Area, in conjunction with the Telemann Festtag in Magdeburg (then in the DDR), and at the Göttingen (Germany) Handel Festival. All of these performances have subsequently been recorded commercially. Works from the Bach database are now in preparation for performance in the Boston area.

(6) Concepts underlying the representations used for searching and analytical tasks, such as Hewlett's base-40 representation for enharmonically-defined intervals, have been incorporated in a number of software products for the teaching of music theory.

There are a great many other foreseeable uses of the data. Programs for instruction in music theory and music appreciation, for rehearsal and performance, for ear-training, for traditional and novel kinds of analysis, for acoustical experimentation and synthetic orchestration, and for research in musical cognition and perception all suggest separate lines of development.

Contents and Sources

The Center's machine-readable files are collected under the names of the various composers whose music has been electronically transcribed. Since notational practices and in consequence representational issues vary from repertory to repertory, the Center's efforts have to date been largely confined to music composed between 1650 and 1825. At this writing (June 1992) no single-composer repertory is yet complete. The Bach DataBase contains approximately three-fourths of the music of J. S. Bach. A few hundred works (in a repertory of a few thousand) each by Handel and Telemann have been transcribed. All of the (72) published works of Corelli are encoded, and one set by the seventeenth-century Giovanni Legrenzi are encoded. Selected sub-repertories by Haydn (quartets), Mozart (quartets), and Beethoven (concertos) are in course of encoding. Miscellaneous works by Vivaldi and Schubert are also encoded.

Rights, Permissions, and Licenses

Input Data

The sources used for CCARH encodings are of four kinds—original manuscripts, early prints, collected editions of the nineteenth century, and newly commissioned unpublished editions. While scholars generally prefer *Gesamtausgaben* (collected editions of the complete works of individual composers) of recent origin for reference and research, performers are less partial, as the production of performance material lags well behind the creation of reference scores and the details do not always correspond closely. The Center has specifically avoided the encoding of such materials in order to avoid any questions of copyright protection. It can thus guarantee that none of the repertory so far encoded is under third-party copyright, with the exception noted below.

Encoded Data

The Center claims copyright on its files to the extent that the selection and organization of the material and the human interpretative skills required to transcribe early notation and to accommodate it to the representation system involve the incessant exercise of informed judgment and human ingenuity. In fact, the representation system provides sufficient latitude that two encoders representing the same work could, under some kinds of circumstances, create files with different contents. Both would be interpretable according to the same general principles.

Third-Party Rights

Third-party rights may pertain to the edition in the case of newly commissioned editions. Some materials encoded directly from manuscript may not be available for publication or performance without the express consent of the owner of the original source. CCARH is unable to anticipate the exact uses to which its data may be put. Users are individually responsible for procuring necessary permissions for publications, performances, and recordings. While making every effort to verify its data, it is unable to warrant that its data will be free of defects.

Issues Related to Access and Acknowledgment

Some issues pertaining to the existence of large corpora of machine-readable sources are common to many projects of an essentially academic character. The Center has had extensive discussions with the U.S. Copyright Office and with other interested parties concerned with analogous issues in the maintenance and use of text databases (*e. g.*, the Packard Humanities Institute, Los Altos, CA; the Center for Electronic Texts in the Humanities, New Brunswick, NJ; the Oxford Text Archive, Oxford, UK; and the study group on rights in electronic networking, Kennedy School of Government, Harvard University) concerning copyright issues pertaining to the creation and use of musical databases. Copyright has been registered on sample *MuseData* files as well as on published editions made from them.

Licensing

The Center intends to make its data available by license. It is anticipated that charges for academic and non-profit use will be minimal. Data may also be licensed to third parties for software development.

Acknowledgment

All users are asked to acknowledge in related software and publications that the musical data has been provided by the Center for Computer Assisted Research in the Humanities.

Eleanor Selfridge-Field

June 18, 1992

CCARH Staff

1992

Walter B. Hewlett, Director (1983—). B.A. (honors), Physics, Harvard University; M.S., Engineering, Stanford University; M.S., Operations Research, Stanford University; D.M.A., Stanford University.

Eleanor Selfridge-Field, Senior Research Associate (1984—). Musicologist. B.A. (honors), Music/history, Drew University; M.S., Journalism, Columbia University; D.Phil., Music, Oxford University.

Edmund Correia, Jr., Music Data Specialist (1985—). Pianist. Scholarship student (German, Music), Stanford University; accompanist, San Jose State University.

Frances Bennion, Music Data Specialist (1985—). Cellist. B.S., Physics/Mathematics, Stanford University.

Steven Rasmussen, Music Data Specialist (1987—). Scholarship student (Music), Princeton University.

Nancy Solomon, Administrative Assistant (1993—). Violinist. B.A., Biology, San Jose State University.

Previous and auxiliary members of the staff have included:

Pamela Decker, Music Data Specialist (1983-85). Organist. D.M.A., Stanford University.
Jean-Pierre Dautrecourt, Music Data Specialist (1985). Composer. Ph.D., Harvard University.
J. Stephen Dydo, Music Programming (1986-7). Composer. D.M.A., Columbia University.
Trudy Brassell, Administrative Assistant (1989-91). B.A., German, Bristol (UK) University.
Michael Flexer, Music Programming (summers, 1985-88). Student (Applied Mathematics), Harvard University (B.A. 1989).
Brent Field, Programming (summers, 1990-92). Student (Cognitive Science), Univ. of California, San Diego.
Meredith Berger, Administrative Assistant (1991-92). B.A., English, Monash University, Victoria, Australia; Dip. Ed., Melbourne University, Victoria, Australia; M.A., Psychology, Santa Clara University.

Technical Documentation

General Principles of Encoding at CCARH

Creation of Logical Records

The creation of logical records in the CCARH databases involves several processes:

(1) Information *about* the music is provided by data entry specialists.

(2) The music is decomposed into tracks, which are played one at a time into a synthesizer.

(3) Each track is stored as a separate file.

(4) A program utilizing the operator information stored in the header records attempts to interpret the music input. It generates a provisional note list on the screen.

(5) The operator corrects obvious errors.

(6) A draft of the track is printed.

(7) Errors are once again corrected.

(8) The corrected file is then stored as part of the database.

Human interpretation is essential to steps (1), (2), (3), (5), and (7). The file that is finally stored in step (8) combines verbal, parametric, logical, and interpreted information. The steps are referred to by number (#) in the following remarks.

1. Verbal Information (#1)

Verbal information stored in the Stage 1 file includes the copyright notice (Records 1 and 2), an identifying work number from a standard reference catalogue (such as BWV for the *Bach-Werke-Verzeichnis*), and the source on which the machine "transcription" is based (Record 3).

2. Parametric Information (#1)

Parametric information is implied by but not necessarily stated in the original score. It includes the following elements of information:

(a) the movement number

(b) the number of sections in the movement

(c) the current section number

(d) the track number

(e) the length of the movement (or section) in measures

(f) the key signature

(g) the number of metronomic units per measure

(h) the number of metronomic units per quarter note

(i) the time signature

(j) the clef

(k) the principal mode

(l) the movement type ("Allegro", "Gigue", etc.)

The items that can be *transferred directly* from the score are (f), (i), (j), and usually (l). Element (l) requires human interpretation when there is no designation in the music.

The items that can be *inferred* from the score are (a), (e), and usually (k). Element (k) requires human interpretation when within the context of a given movement, the mode is variable and when the harmonic usage is at variance with the key signature.

The remaining items, in common with (k) and (l), require contextual understanding and are therefore considered to be contextual data attributes. An extensive list of contextual data attributes is given in Table 1.

3. Logical Information (#4, #6, #8)

Logical information is assembled by the computer in step (4), utilized in step (6), and stored in step (8). Pitch name, enharmonic inflection name, octave number, and duration are treated as logical information. These items of information are stored in the note-list (the main body of the file).

4. Interpreted Information (I): Legibility and Basic Meaning

There are many dialects of Common Musical Notation. Because of the Center's orientation toward historical repertories and source preservation, its work is enmeshed in notational traditions that were somewhat less precise, less logical, and less consistent than that in use today. In particular, there were many conventions used in earlier centuries that were commonly understood to mean something different aurally from what the writing on the page suggests to today's performers. Some specific areas of

imprecision and ambiguity involve rhythmic, melodic, harmonic, and accompaniment conventions and are itemized in Table 1.

More mundane matters such as crowding of notes, fading of notation, and obliteration of notation by mildew and other agents of deterioration impede the simple task of reading. Some examples of the eighteenth-century prints and manuscripts from which we read are included as illustrations 1, 2, 3, and 4. Such sources can only be accommodated to machine readability by substantial amounts of human intervention.

5. Interpreted Information (II): Contextual Data Attributes (#2, #5, #7)

Contextual data attributes are those attributes that can only be defined by the context in which they occur. The context may be limited to one of the three realms of musical apperception (logical, graphic, acoustical). Timbre is an example of an acoustical attribute that has no graphic representation. Accent is an example of a logical attribute that has no graphic representation. What often makes these attributes noticeable is a conflict of needs between one of them and (a) the non-logical aspects of common musical notation or (b) the arbitrary nature of the data entry process.

A simple example of a logical data attribute that can only be understood contextually is the dot: if it is on the same horizontal axis as the note preceding it, it formally extends duration, but if it is on the same vertical axis, it informally reduces duration.

The dot is also an apt illustration of problem type (a), since in earlier times its meaning (on the horizontal axis) was variable: it elongated the note that preceded it by an amount that had to be inferred from the context. Some uncertainty about precise duration lingers. Thus any effort to give it a precise meaning today is an act of interpretation.

Although parts are entered one at a time, some aspects of the relationship between parts must be indicated in the representation. Some rhythmic notations were commonly understood to be played differently from what their appearance would suggest. A large number of specific types of contradictions have been codified by CCARH to facilitate the process of consistency in interpretation. (See Illustrations 5 and 6.) Some contextual considerations that are imposed by the data entry process but must be resolved by reference to the original source are (b), (c), (d), (g), and (h).

Beyond this, the data entry process imposes a need for precision where no precise method of musical notation has ever been adopted. Some recurrent situations that require redefinition are listed in Table 1.

Contextual Data Attributes

A. Related to barring
   1. Non-complementarity of first and last bars
   2. Non-consistency between parts of first and last bar representation
   3. Bar count discrepancy between parts
   4. Bar count discrepancy in treatment of hemiolas

B. Related to movement definitions
   1. Contradiction in use of single and double barlines
   2. Contradiction in use of repeat signs
   3. Contradiction between tempo and character designations
   4. Contradictions caused by succession of meters coupled with one theme

C. Related to rhythm
   1. Time signature does not correspond to barring
   2. Temporary changes in subdivision of bar or beat
   3. Implied differences in time signatures between parts
   4. Ambiguity between ties and slurs
   5. Need to accommodate unusual subdivisions of the beat
   6. Rhythmic notation illogical
   7. Duple/triple contradictions in vertical relationship of parts (2:3, 3:4, 4:6, etc.)

D. Related to pitch
   1. Discrepancies between "duplicate" parts
   2. Misplaced accidentals
   3. Wrong sign for accidental
   4. Octave discrepancy
   5. Misplaced basso continuo figures
   6. Inconsistent use of accidentals
   7. Uncertain duration of accidental
   8. Part incomplete; missing bars deducible from surrounding material

Table 1. Contextual Data Attributes

6. Interpreted Information (III): Error Correction (#2, #5, #7)

While in general the aim is to provide a faithful transcription of the original source, some categories of errors and omissions must be corrected in order to permit logical organization and sequential processing. In addition, the resolution of the kinds of problems listed under contextual data attributes is essential. (See Illustration 7.)

Organization of Logical Records

The most far-reaching aspect of contextual interpretation is the arbitrary decomposition of an integrated musical work into tracks. For most vocal and instrumental parts, a track is indistinguishable from a part. Variations in texture (*e.g.*, chords in a single-line melody) producing oddities of coincidence and non-coincidence (represented in traditional scores by aurally concurrent notes that are not quite concurrent graphically) must be handled on a case-by-case basis. Part crossings are not always easily determined. Arpeggiation and other free-style figuration present especially difficult programs of representation.

In addition, logical tracks may be required where graphic tracks are only implied. "Tutti" cues imply the existence of parts/tracks that are not shown on the page. Some recurrent problems that require human resolution are these:

1. Number of tracks is ambiguous (Illustrations 8a, b, c) on account of
   A. Dual-stem notation
   B. Absence or inconsistent use of rests
   C. Staggering of inner voices
   D. Contradictions between stem direction and thematic sense
   E. *Style brisé* (*Freistimmigkeit*)
   F. Inconsistencies of texture

2. Path of track is ambiguous (Illustration 9)

3. Derivation of multiple tracks from one part unclearly specified, as in
   A. Solo/tutti cues applied to single part

B. Contradictory cues for entry and exit of instruments

C. Ranges that exceed limits of specified instrument

D. Composite basso continuo parts (one part, several instruments)

In consequence of these considerations, the number of tracks required to represent a piece of music is far greater than one might expect. The number, the content of each, and the sequence in which they are arranged are unlikely to be duplicated in any other system, are not always replicated within our own system, and were certainly never formulated by the composers or any editors making editions of their music. The organization of logical records is thus 100% dependent on human interpretation.

Two brief examples may serve to illustrate the unpredictable dimensions taken on by the data as it is organized into databases.

Bach's fourth orchestral suite (BWV 1069) occupies 67 files. There are eleven movements. The work is scored for six ensemble instruments and basso continuo (seven parts minimum). The scoring of individual movement is variously for three, four, and six instruments and variously with and without continuo accompaniment.

The Ouverture of Telemann's *Musique de Table* occupies 90 files. The work has ten movements and there are nine tracks in each movement. A printout of the files containing this one work runs to 633 pages; a printout of the music itself, track by track, is 126 pages.

While the decomposition of a score into tracks may take many courses, the correct reconstruction of a score from tracks is assured by the provision of a statement in the index records as to the total number of tracks for any given movement.

Illustration 1. Eighteenth-century print (good quality)

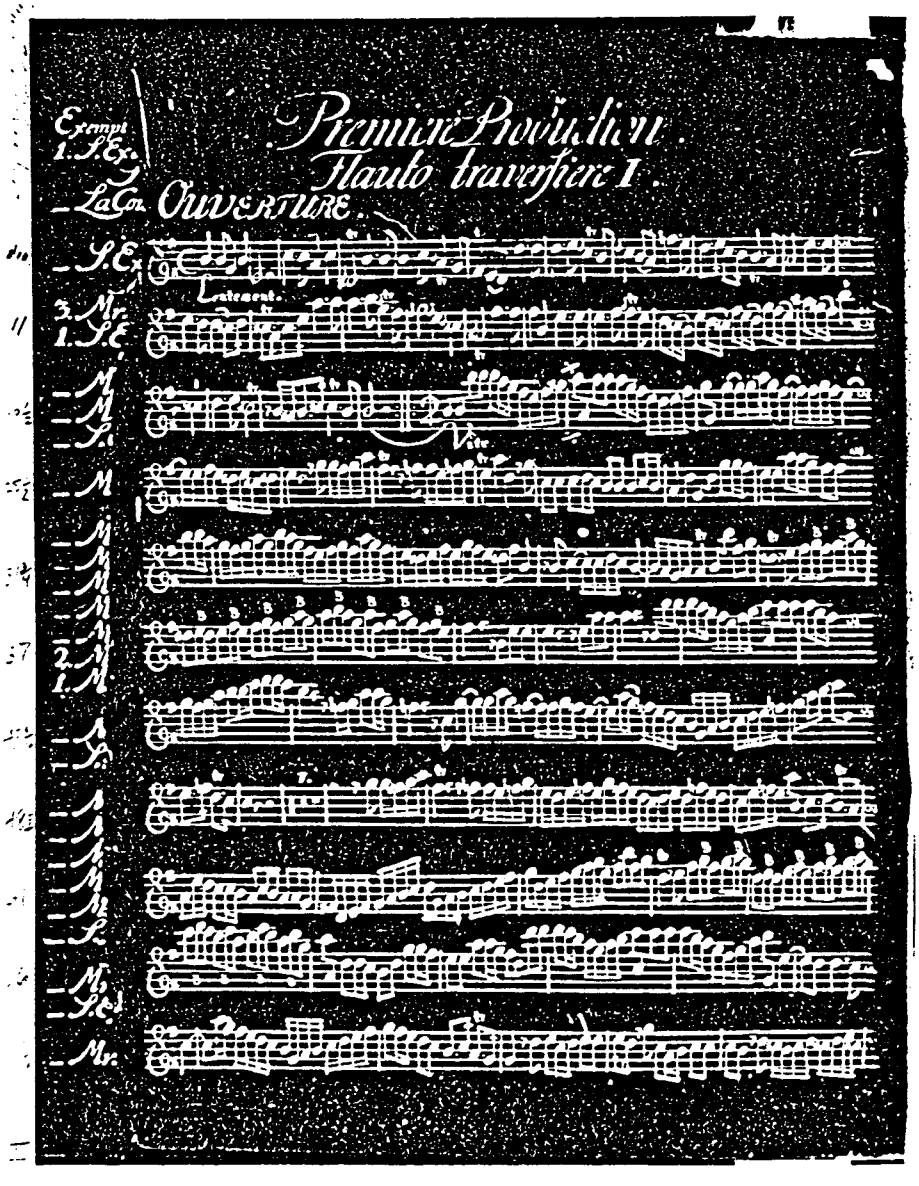
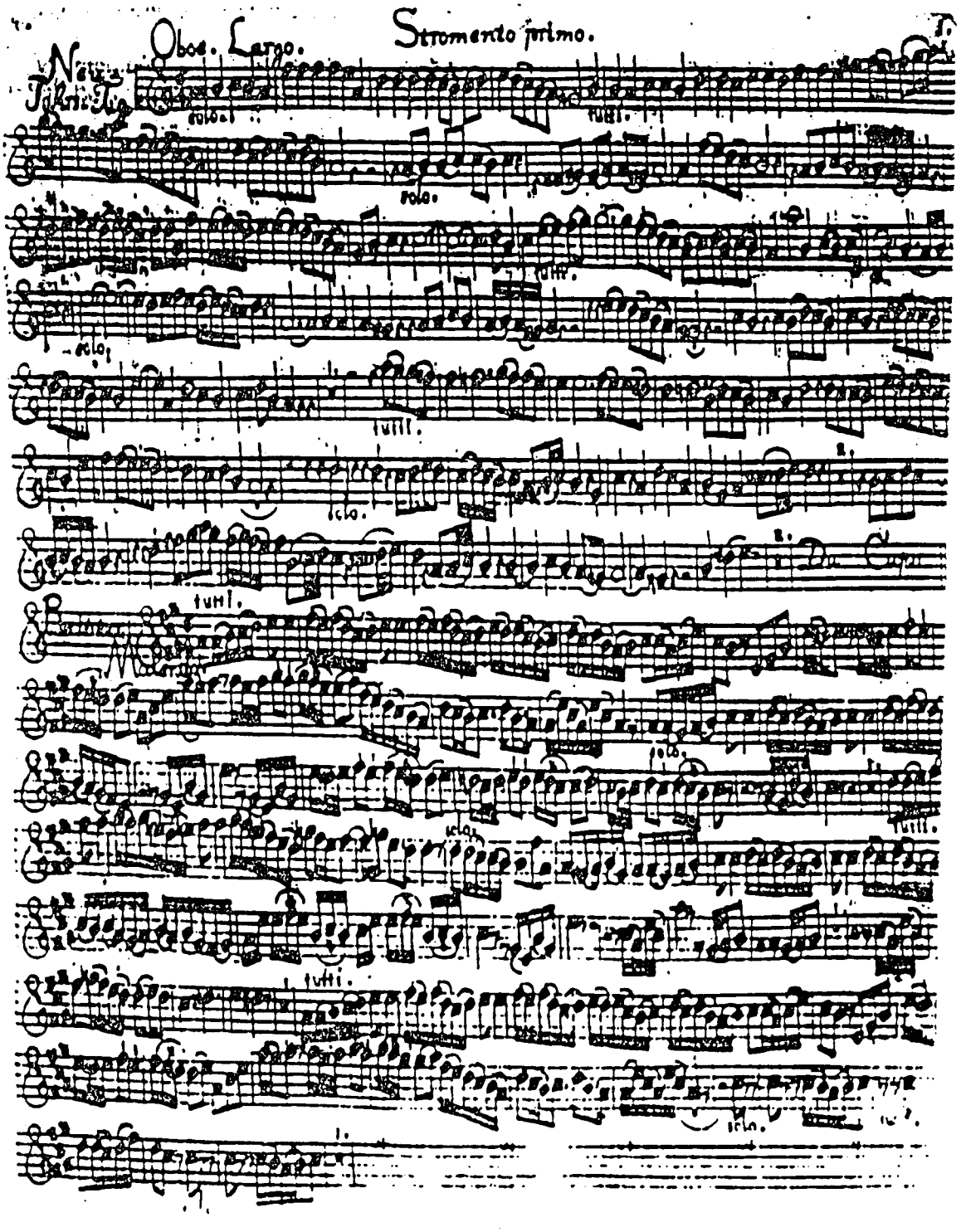
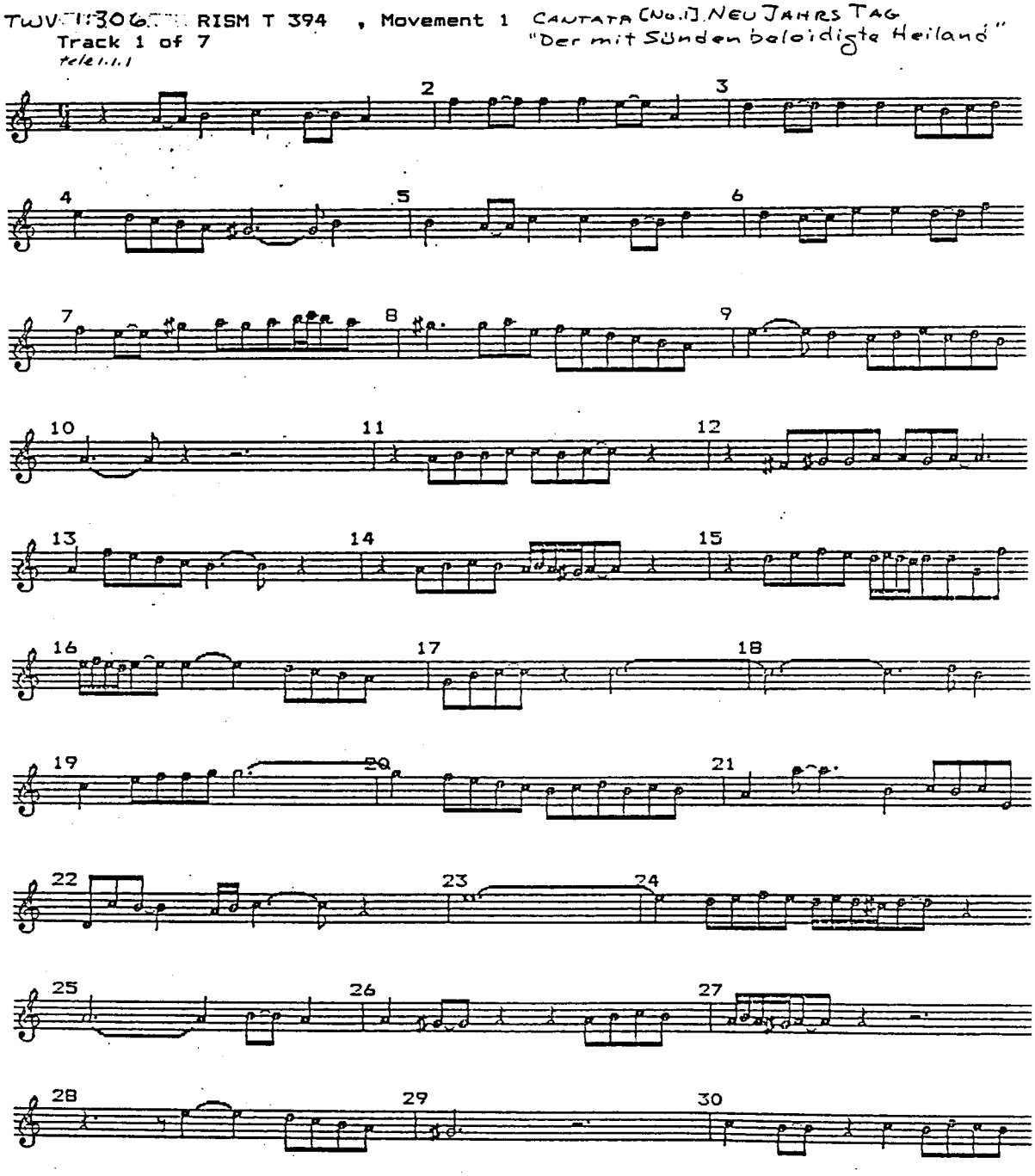


Illustration 2. Eighteenth-century print (poor quality) and its handling

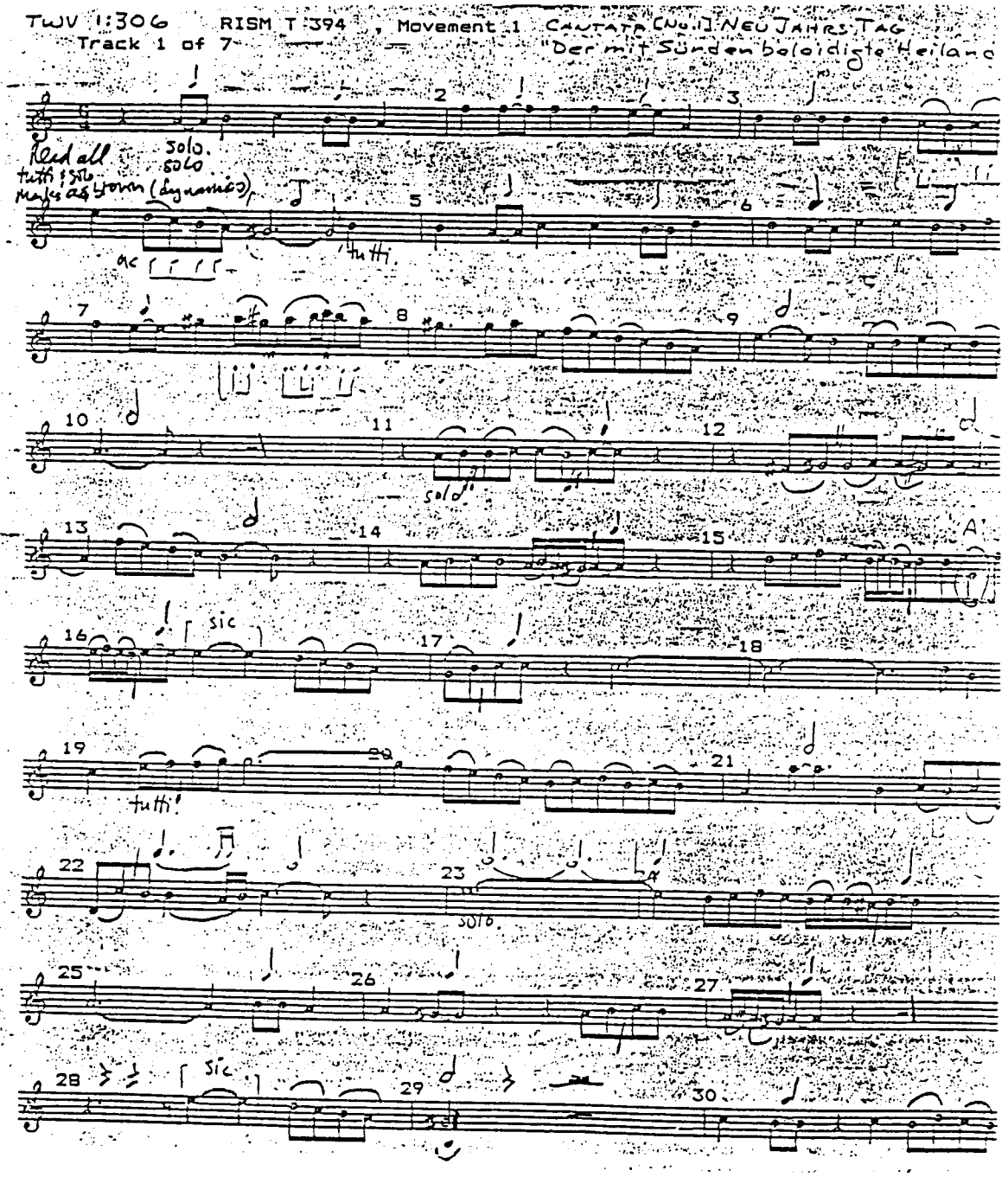
a. Original print



b. Step #6


TWV 1:306 RISM T 394 , Movement 1 CANTATA [No.1] NEU JAHRS TAG
Track 1 of 7
"Der mit Sünden beleidigte Heiland"
tele1.1.1

c. Step #7


TWV 1:306 RISM T 394, Movement 1 CANTATA (No. 1 NEUJAHRS TAG
Track 1 of 7
"Der mit Sünden beloidigte Heilano
Read all
Solo.
Solo
tutti / Solo
Marks as Horn (dynamics)
tutti
tutti
Solo.
sic
tutti!
Solo.
sic

d. Step #8

```
447          Number of records
1:306 1      TWV Number and Movement
DK-Kk mu 651 )
0.0535 u260  } Source
RISM T 394   )
7 1          Total No. of tracks and current track
59 0 (48 8)  Length in measures, Key, Beeps/meas, Beeps/quarter note
6 4, 0 0 75  Time signature (4/4), Clef, mode, type
measure 1
rest    8
A4      8
B4      8
C5      8
B4      8
A4      8
measure 2
F5      8
F5      8
F5      8
F5      8
E5      8
A4      8
measure 3
D5      8
D5      8
D5      8
D5      8
C5      4
B4      4
C5      4
D5      4
measure 4
E5      8
D5      4
C5      4.
B4      4
A4      4
G#4     16
B4      8
measure 5
B4      8
A4      8
C5      8
C5      8
B4      8
D5      8
measure 6
D5      8
C5      8
E5      8
E5      8
D5.     8
F5      8
measure 7
F5      8
```

Clef 0 = treble
Clef 1 = bass

Mode 0 = minor
Mode 1 = major
Mode 2 = minor-[illegible] form flat in key [illegible] (count the 2 flats in key signature)

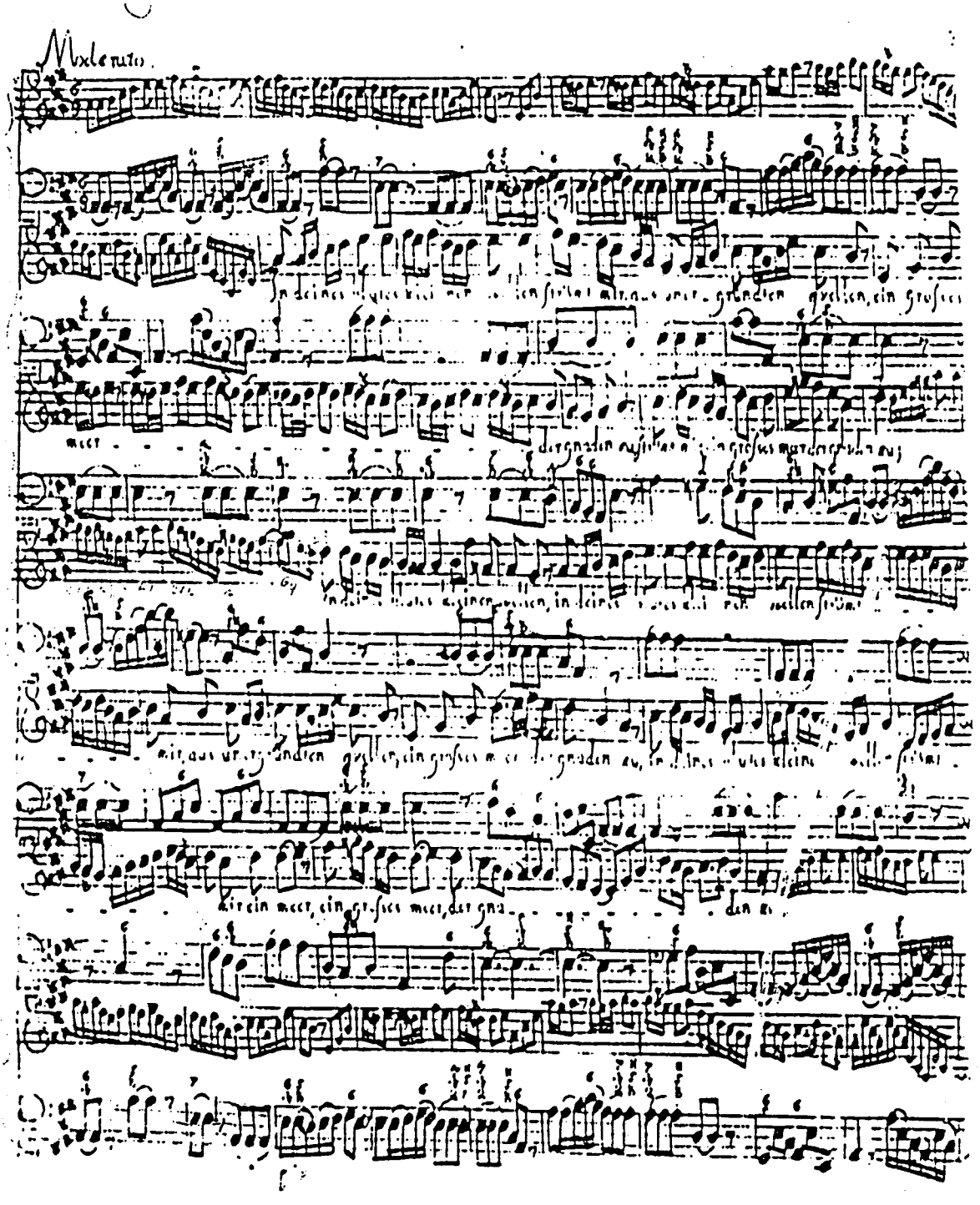
Type 75 = Cantata - Aria (1 [illegible])
Type 79 = Cantata - Recitative

"-" after duration = tie

* and !* = our editorial [illegible]
* = an assumed accid[ental] (not in manuscript), discrepancy [illegible]
!* = [illegible] [illegible] when there is an accidental on the [illegible] note.

Illustration 3. Eighteenth-century manuscript (good quality)

a. Original manuscript



b. Clarification of questions of pitch, rhythm, and barring

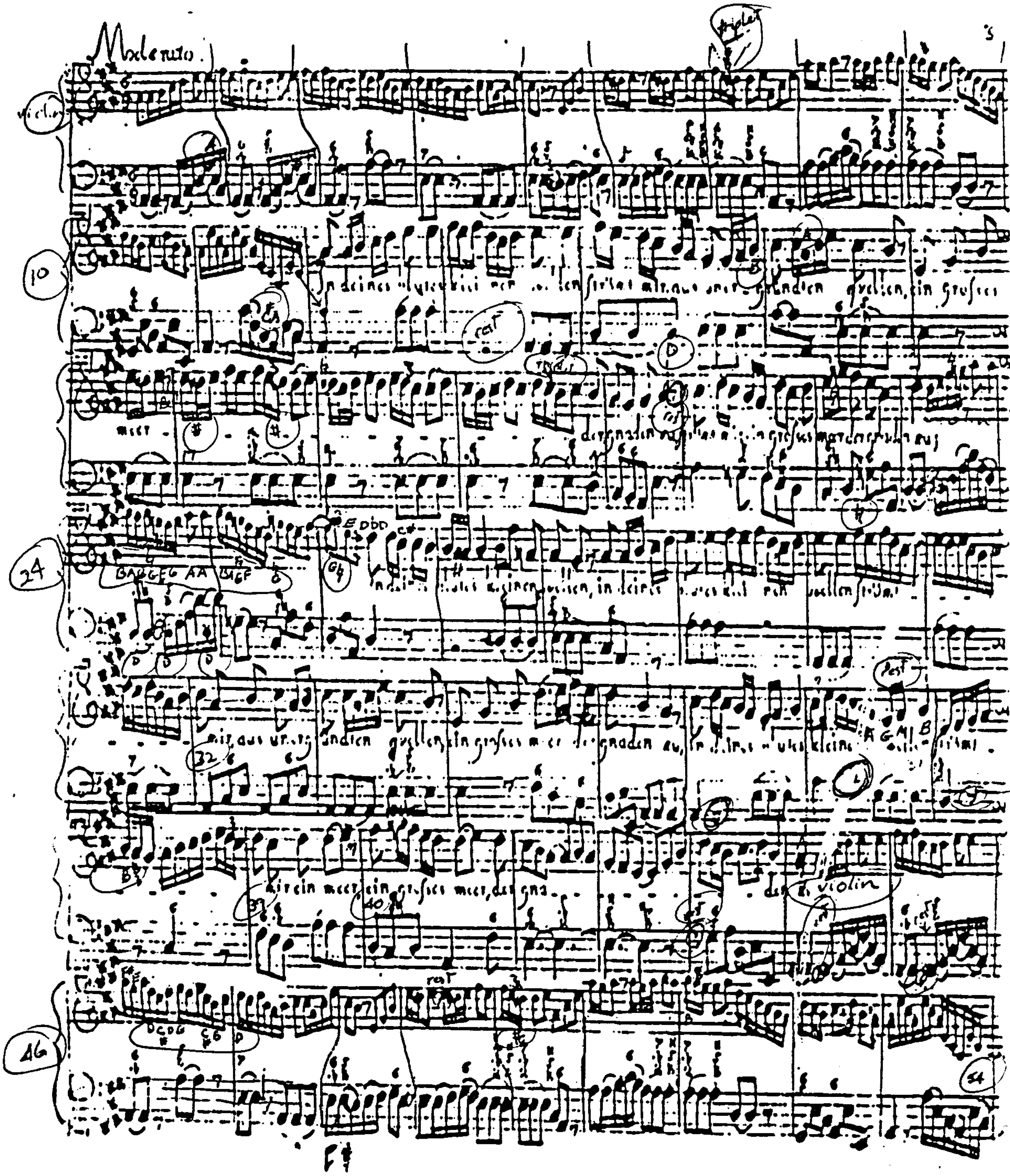


Illustration 4. Eighteenth-century manuscript (poor quality)

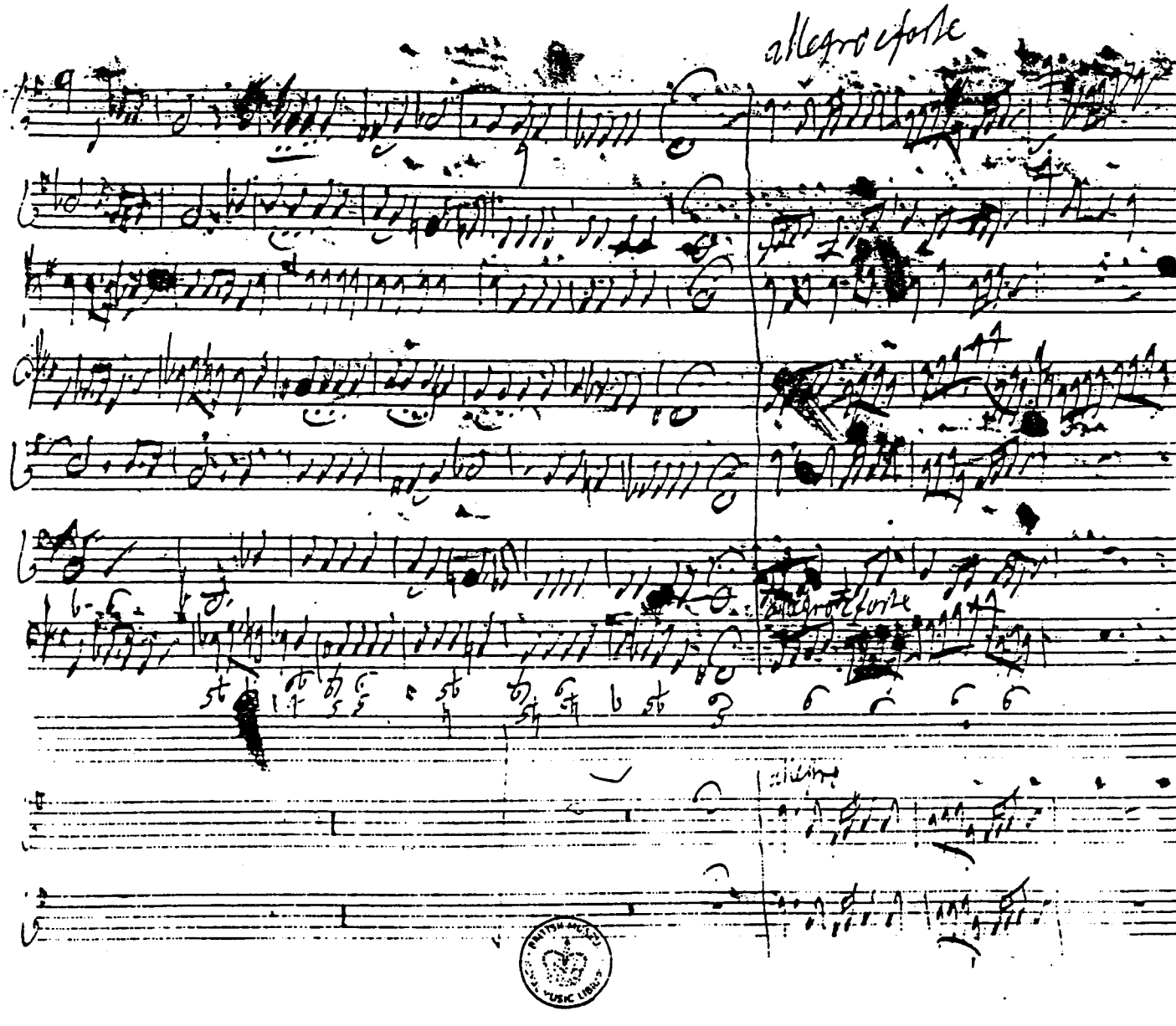


Illustration 5. Contextual attributes: part realignment

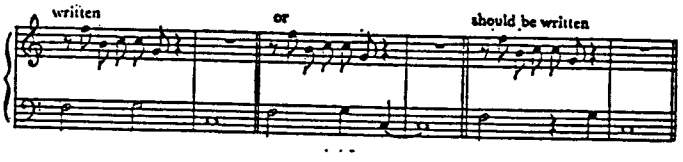


Illustration 6. Contextual attributes: need for rhythmic interpretation

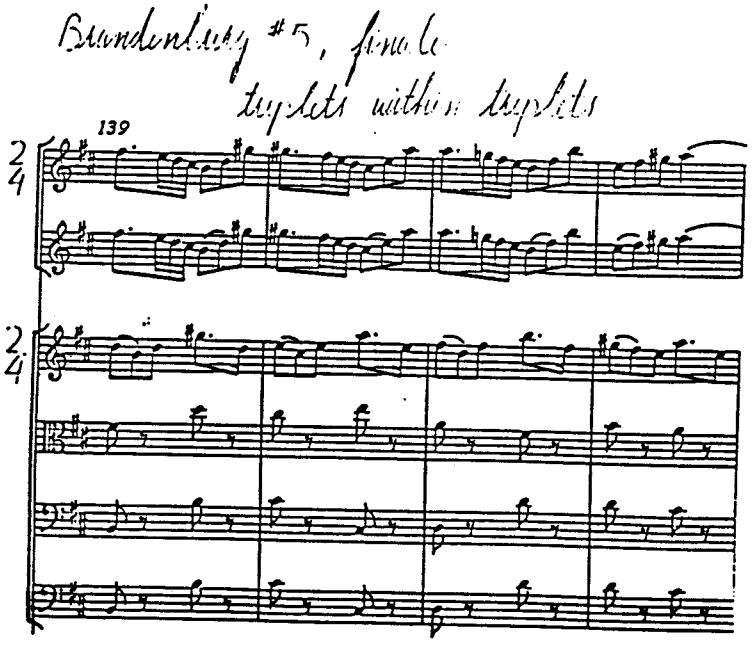
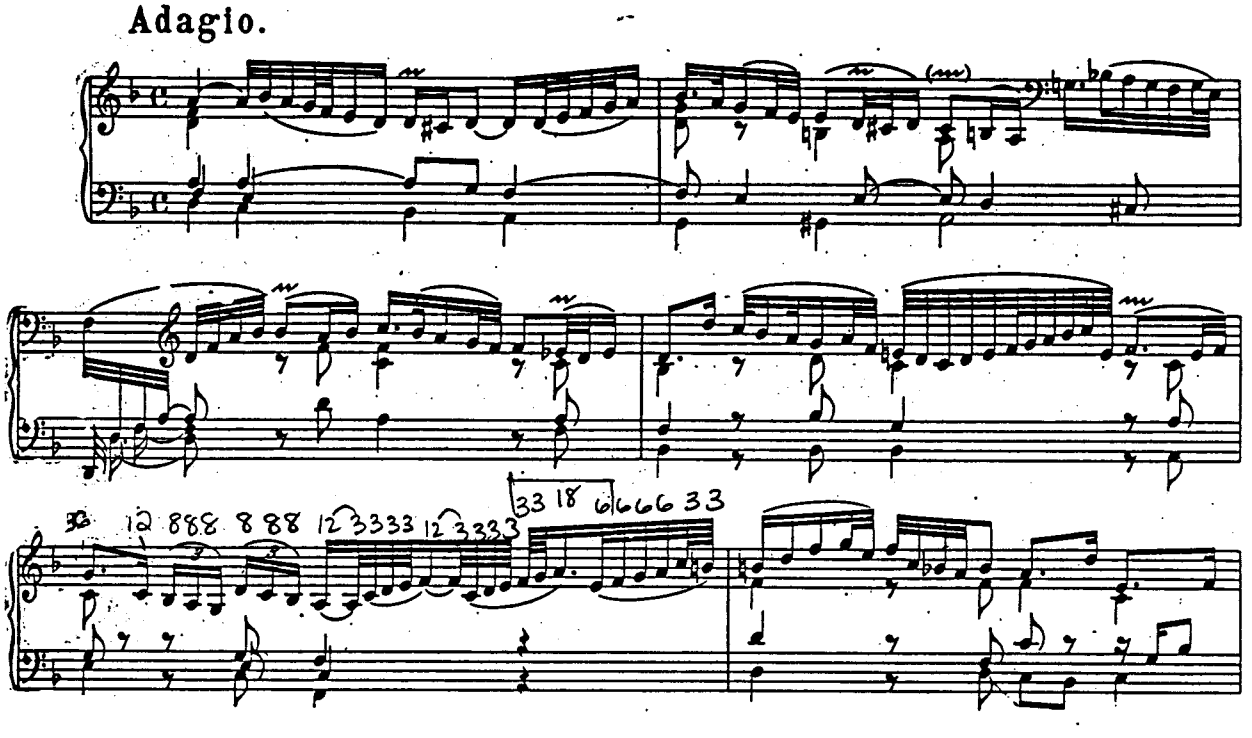


Illustration 7. Error correction: embedded error in complex rhythm

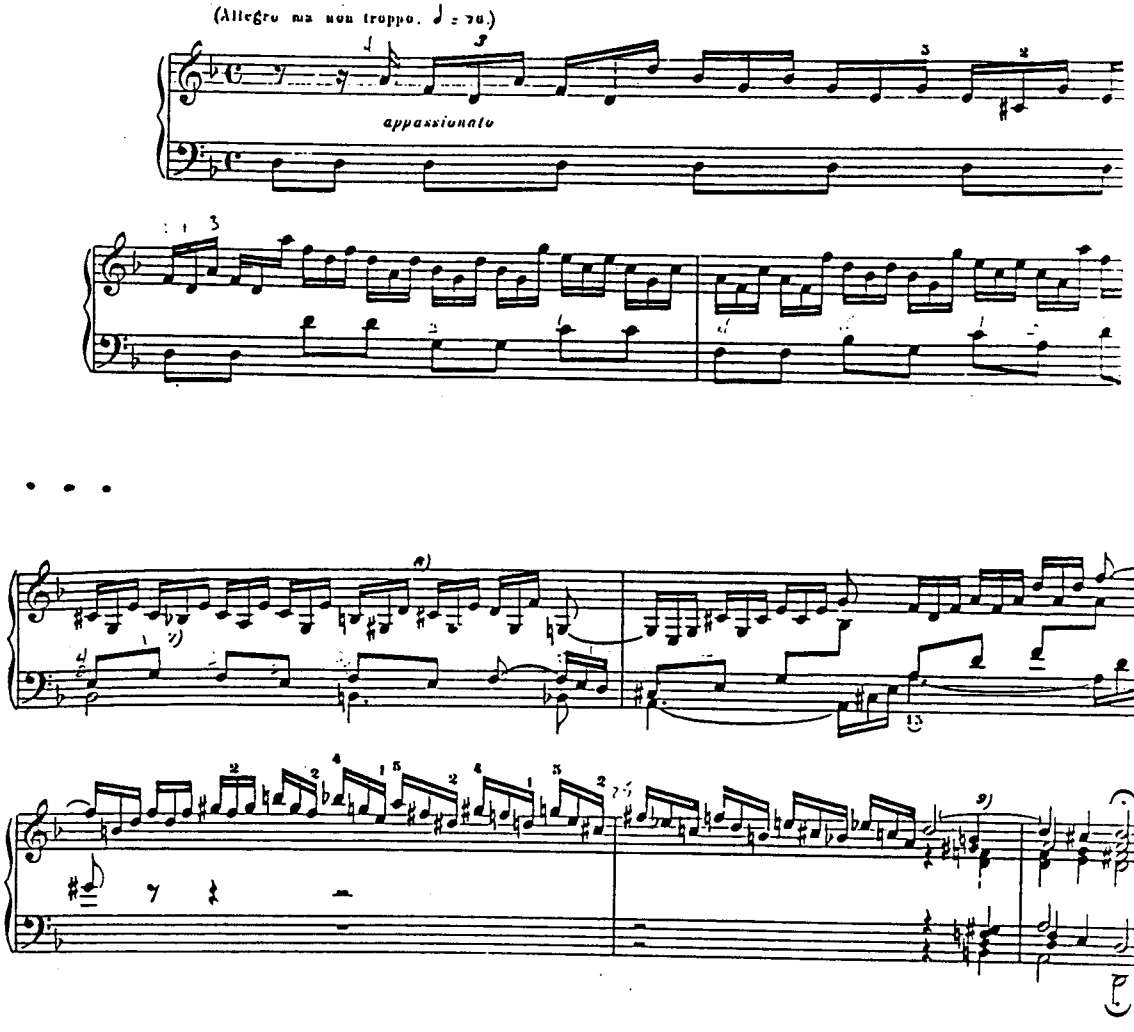



```
1     116
2     4444  444
3     Bach Gesells
4     chaft xix
5
6     6 1
7     5 -1 192 48
8     4 4 0
9     measure 1
10    A4    54
11    Bf4    6
12    A4     6
13    G4     3
14    F4     3
15    E4    12
16    D4    12
17    D4    12
18    C#4   12
19    D4    36
20    D4     6
21    E4     6
22    F4     6
23    G4     6
24    A4    12
25    measure 2
26    Bf4   18
27    A4     6
28    G4    12
29    F4     6
30    E4     6
31    E4    24
32    D4     6
33    C#4    6
34    D4    12
35    C#4   24
36    B3    12
37    A3    12
38    G3    18
39    Bf3    6
40    A3     6
41    G3     6
42    F3     6
43    G3     3
44    E3     3
45    measure 3
46    F3     6
47    D3     6
48    F3     6
49    A3     6
50    D4     6
51    F4     6
52    A4     6
53    Bf4    6
54    Bf4   24
55    A4    12
56    Bf4   12
57    C5    18
58    Bf4    6
59    A4    12
60    G4     6
61    F4     6
62    F4    24
63    Ef4    6
64    D4     6
65    Ef4   12
66    measure 4
67    D4    36
68    D5    12
69    C5     6
70    Bf4   12
71    A4     6
72    G4    12
73    A4     6
74    F4     6
75    E4     6
76    D4     6
77    C4     6
78    D4     6
79    E4     6
80    F4     3
81    G4     3
82    A4     3
83    B4     3
84    C5     3
85    deleted
86    E4     3
87    F4    36
88    E4     6
89    F4     6
90    measure 5
91    G4    36
92    C4    12
93    Bf3    8
94    A3     8
95    G3     8
96    D4     8
97    C4     8
98    Bf3    8
99    A3    15
100   C4     3
101   D4     3
102   E4     3
103   F4    15
104   C4     3
105   D4     3
106   E4     3
107   F4     3
108   G4     3
109   A4    12
110   E4     6
111   F4     6
112   G4     6
113   A4     6
114   C5     3
115   D4     3
116   deleted
117   END
```

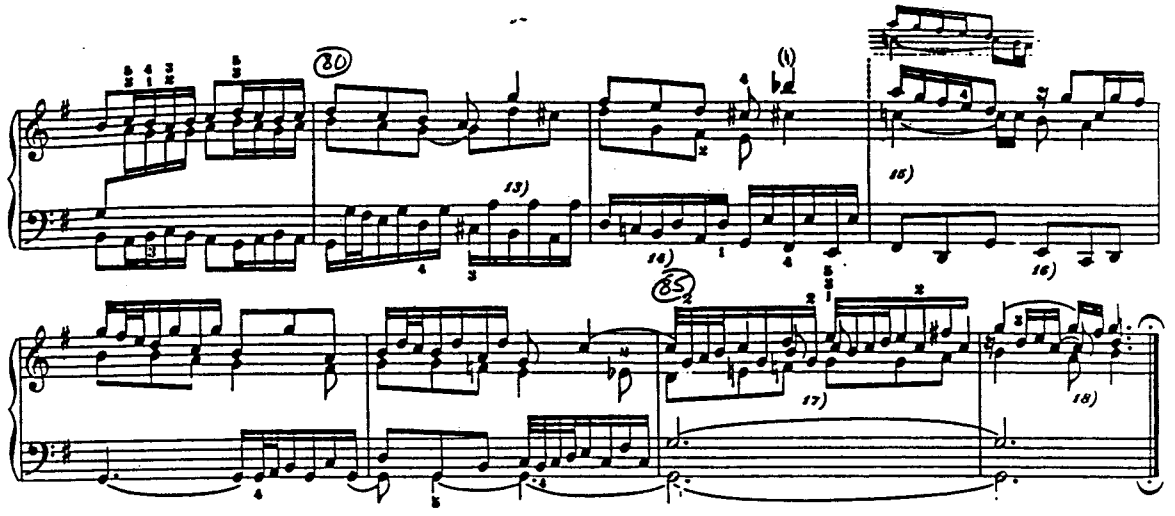
The representation of measure 5 is divided into quarters. The fourth quarter has an overflow value of 54 where one would expect 48.

Illustration 8. Organization of logical records: number of tracks ambiguous

a. Two tracks expanded to six



b. Three tracks expanded to five



c. Two tracks expanded to four

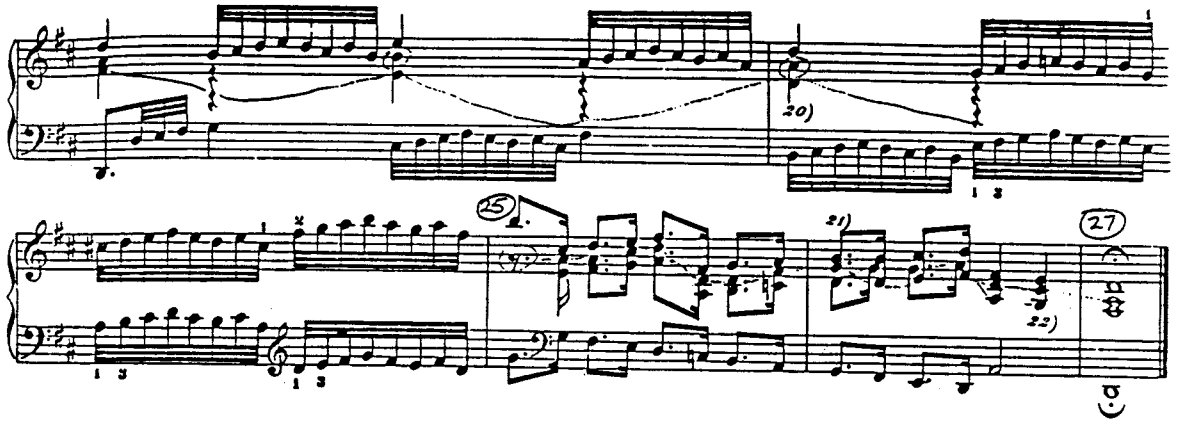
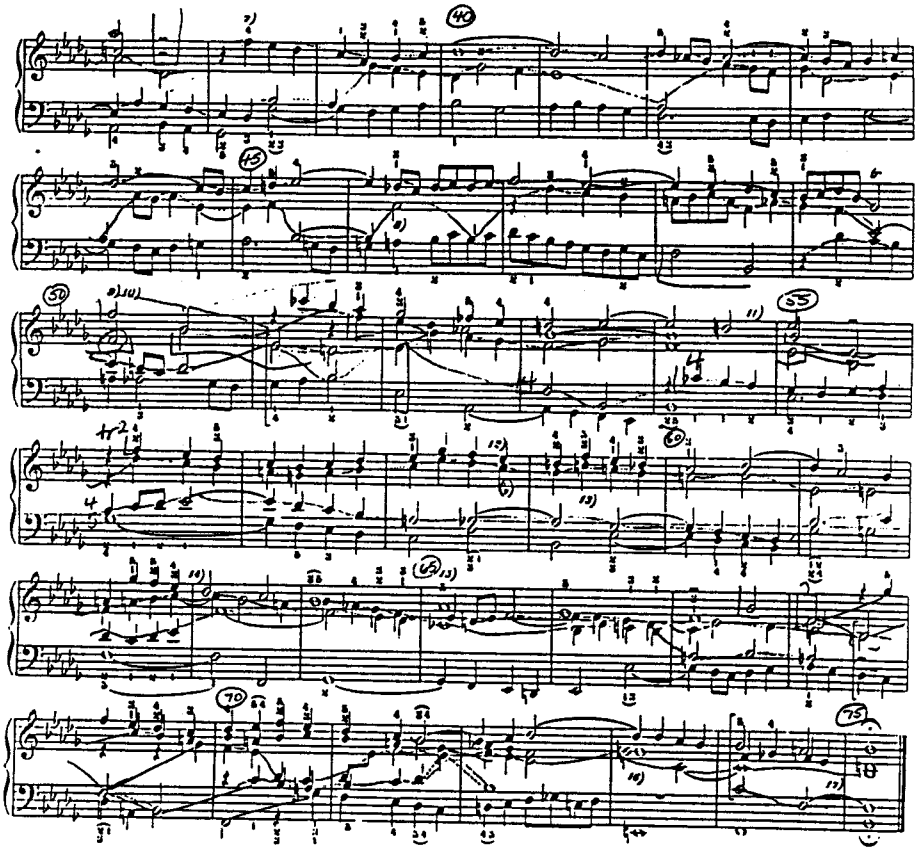


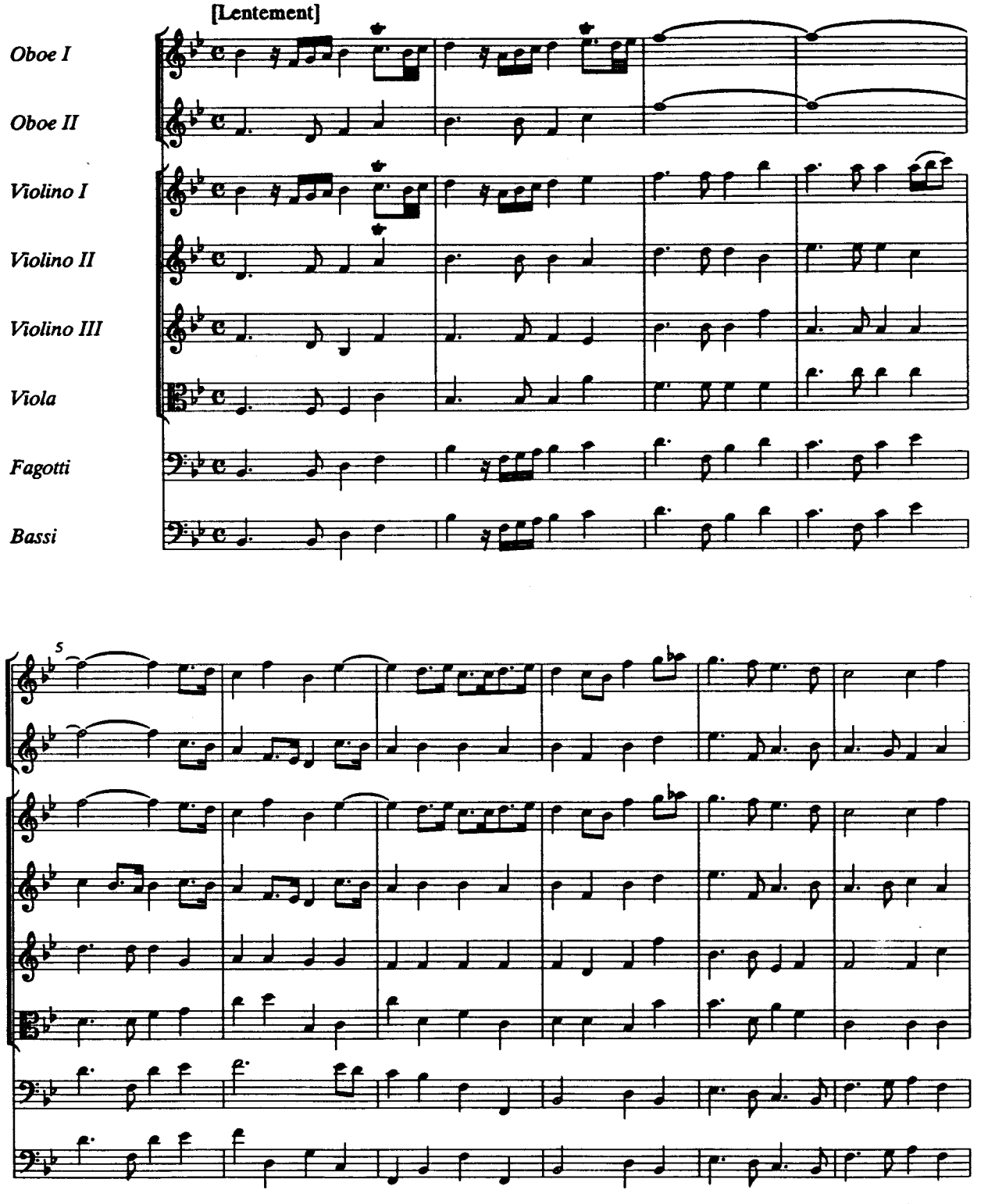
Illustration 9. Organization of logical records: track paths ambiguous



Sound Information

(Stage 1)

Sample Score: G. F. Handel, *Ottone*, Overture, Full Score, Bars 1 - 10



Input Files for Parts: G. F. Handel, *Ottone*, Parts 1 - 4, Bars 1 - 10 (Stage 1 encoding)

```
107
15     0,1
Hicks/Chry.
1991

8 1
16 -2 32 8
4 4 0 1 18
measure 1
Bf4    8
rest   2
F4     2
G4     2
A4     2
Bf4    8
C5     6
Bf4    1
C5     1
measure 2
D5     8
rest   2
A4     2
Bf4    2
C5     2
D5     8
Ef5    6
D5     1
Ef5    1
measure 3
F5    32-
measure 4
F5    32-
measure 5
F5    16-
F5     8
Ef5    6
D5     2
measure 6
C5     8
F5     8
Bf4    8
Ef5    8-
measure 7
Ef5    8
D5     6
Ef5    2
C5     6
C5     2
D5     6
Ef5    2
measure 8
D5     8
C5     4
Bf4    4
F5     8
G5     4
Af5    4
measure 9
G5    12
F5     4
Ef5   12
D5     4
measure 10
C5    16
C5     8
F5     8
measure 11
```

```
81
15     0,1
Hicks/Chry.
1991

8 2
16 -2 16 4
4 4 0 1 18
measure 1
F4     6
D4     2
F4     4
A4     4
measure 2
Bf4    6
Bf4    2
F4     4
C5     4
measure 3
F5    16-
measure 4
F5    16-
measure 5
F5     8-
F5     4
C5     3
Bf4    1
measure 6
A4     4
F4     3
Ef4    1
D4     4
C5     3
Bf4    1
measure 7
A4     4
Bf4    4
Bf4    4
A4     4
measure 8
Bf4    4
F4     4
Bf4    4
D5     4
measure 9
Ef5    6
F4     2
A4     6
Bf4    2
measure 10
A4     6
G4     2
F4     4
A4     4
measure 11
```

```
113
15     0,1
Hicks/Chry.
1991

8 3
16 -2 32 8
4 4 0 1 18
measure 1
Bf4    8
rest   2
F4     2
G4     2
A4     2
Bf4    8
C5     6
Bf4    1
C5     1
measure 2
D5     8
rest   2
A4     2
Bf4    2
C5     2
D5     8
Ef5    8
measure 3
F5    12
F5     4
F5     8
Bf5    8
measure 4
A5    12
A5     4
A5     8
A5     2
Bf5    2
C6     4
measure 5
F5    16-
Ef5    8
Ef5    6
D5     2
measure 6
C5     8
F5     8
Bf4    8
Ef5    8-
measure 7
Ef5    8
D5     6
Ef5    2
C5     6
C5     2
D5     6
Ef5    2
measure 8
D5     8
C5     4
Bf4    4
F5     8
G5     4
Af5    4
measure 9
G5    12
F5     4
Ef5   12
D5     4
measure 10
C5    16
C5     8
F5     8
measure 11
```

```
91
15     0,1
Hicks/Chry.
1991

8 4
16 -2 16 4
4 4 0 1 18
measure 1
D4     6
F4     2
F4     4
A4     4
measure 2
Bf4    6
Bf4    2
Bf4    4
A4     4
measure 3
D5     6
D5     2
D5     4
Bf4    4
measure 4
Ef5    6
Ef5    2
Ef5    4
C5     4
measure 5
C5     4
Bf4    3
A4     1
Bf4    4
C5     3
Bf4    1
measure 6
A4     4
F4     3
Ef4    1
D4     4
C5     3
Bf4    1
measure 7
A4     4
Bf4    4
Bf4    4
A4     4
measure 8
Bf4    8
F4     8
Bf4    8
D5     8
measure 9
Ef5    6
F4     2
A4     6
Bf4    2
measure 10
A4     6
Bf4    2
C5     4
A4     4
measure 11
```

G. F. Handel, *Ottone*, Parts 5 - 8, Bars 1 - 10 (Stage 1 encoding)

```
90
15     0,1
Hicks/Chry.
1991

8 5
16 -2 16 4
4 4 0 1 18
measure 1
F4     6
D4     2
Bf3    4
F4     4
measure 2
F4     6
F4     2
F4     4
Ef4    4
measure 3
Bf4    6
Bf4    2
Bf4    4
F5     4
measure 4
A4     6
A4     2
A4     4
A4     4
measure 5
D5     6
D5     2
D5     4
G4     4
measure 6
A4     4
A4     4
G4     4
G4     4
measure 7
F4     4
F4     4
F4     4
F4     4
measure 8
F4     4
D4     4
F4     4
F5     4
measure 9
Bf4    6
Bf4    2
Ef4    4
F4     4
measure 10
F4     8
F4     4
C5     4
measure 11
```

```
87
15     0,1
Hicks/Chry.
1991

8 6
16 -2 16 4
4 4 13 1 18
measure 1
F3     6
F3     2
F3     4
C4     4
measure 2
Bf3    6
Bf3    2
Bf3    4
A4     4
measure 3
F4     6
F4     2
F4     4
F4     4
measure 4
C5     6
C5     2
C5     4
C5     4
measure 5
D4     6
D4     2
F4     4
G4     4
measure 6
C5     4
D5     4
Bf3    4
C4     4
measure 7
C5     4
D4     4
F4     4
C4     4
measure 8
D4     4
D4     4
Bf3    4
Bf4    4
measure 9
Bf4    6
D4     2
A4     4
F4     4
measure 10
C4     8
C4     4
C4     4
measure 11
```

```
91
15     0,1
Hicks/Chry.
1991

8 7
16 -2 16 4
4 4 24 1 18
measure 1
Bf2    6
Bf2    2
D3     4
F3     4
measure 2
Bf3    4
rest   1
F3     1
G3     1
A3     1
Bf3    4
C4     4
measure 3
D4     6
F3     2
Bf3    4
D4     4
measure 4
C4     6
F3     2
C4     4
Ef4    4
measure 5
D4     6
F3     2
D4     4
Ef4    4
measure 6
F4    12
Ef4    2
D4     2
measure 7
C4     4
Bf3    4
F3     4
F2     4
measure 8
Bf2    8
D3     4
Bf2    4
measure 9
Ef3    6
D3     2
C3     6
Bf2    2
measure 10
F3     6
G3     2
A3     4
F3     4
measure 11
```

```
92
15     0,1
Hicks/Chry.
1991

8 8
16 -2 16 4
4 4 24 1 18
measure 1
Bf2    6
Bf2    2
D3     4
F3     4
measure 2
Bf3    4
rest   1
F3     1
G3     1
A3     1
Bf3    4
C4     4
measure 3
D4     6
F3     2
Bf3    4
D4     4
measure 4
C4     6
F3     2
C4     4
Ef4    4
measure 5
D4     6
F3     2
D4     4
Ef4    4
measure 6
F4     4
D3     4
G3     4
C3     4
measure 7
F2     4
Bf2    4
F3     4
F2     4
measure 8
Bf2    8
D3     4
Bf2    4
measure 9
Ef3    6
D3     2
C3     6
Bf2    2
measure 10
F3     6
G3     2
A3     4
F3     4
measure 11
```

Technical Description

Sound (Stage 1) File Format

DataBase (Bach) Format

(October 2, 1984; rev. December 5, 1984)

[This description was written for the Stage 1 encodings of Bach's *Well-Tempered Clavier*, which are based on the Bischoff edition (1888) and were deposited in 1986 in the Oxford Text Archive.]

The record length is 256 bytes, organized as follows:

Header record (Record 0):

Bytes 1-24:	Composer
Bytes 25-72:	Work
Bytes 73-120:	Source
Bytes 121-122:	Number of data groups

Index Records (Records 1 through 10):

Each movement will have a three-word reference in the index:

Word 1: BWV number

Bits 0-3:	Version of BWV number (0 = normal)
Bits 4-15:	BWV number

Word 2: Movement number

Bits 0-2:	Version of movement (0 = normal)
Bits 3-15:	Movement number

Word 3: Record number in the file where this movement starts

Data Records (Records 11...):

Each movement will start a new record. The first three words in this data record are as follows:

Word 1:

Bits 0-3: Version of BWV number (0 = normal)

Bits 4-15: BWV number

Word 2:

Bits 0-2: Version of movement (0 = normal)

Bits 3-8: Section of movement (0 = normal)

If more than one section:

Bits 3-5: Section number

Bits 6-8: Number of sections

Bits 9-15: Movement number

Word 3: Number of records for this movement (section)

The following 12 words will give information about the movement (section):

Word 4:

Byte 1: Number of divisions per quarter note

Byte 2: Number of divisions per measure (0—256)

Word 5: Number of measures

Word 6: Number of tracks (n = number of tracks)

Word 7: Key (-7 [seven flats] to +7 [seven sharps])

Word 8: Time signature

Byte 1: Time numerator

Byte 2: Time denominator

Word 9: Mode

1 = Major

0 = Minor

2 = Dorian (minor; missing one flat)

3 = Mixolydian (major; missing one sharp)

4 = Phrygian (minor; missing one sharp)

5 = Lydian (major; missing one flat)

6 = Double Dorian (minor; missing two sharps)

Word 10: Movement type (more than 80 in use in 1992)

Word 11: First measure number (1 = normal; 0 = pickup; >1 = new section)

Word 12: Reserved

Word 13: Reserved

Word 14: Reserved

Word 15: Reserved

The following "n" words contain pointers to the beginning of the (n) tracks. These are given as byte addresses, starting with the first byte of the first movement record. Following these pointers are the data for the tracks themselves. The structure of the "note" word is as follows:

Note Word:

Byte 1:

```
  0 = rest
1-40 = C1ff-B1##
41-80 = C2ff-B2##
81-120 = C3ff-B3##
121-160 = C4ff-B4##
161-200 = C5ff-B5##
201-240 = C6ff-B6##
241-245 = C7ff-C7##
  246 = 2-note flag
  247 = 3-note flag
  248 = 4-note flag
  249 = 5-note flag
  250 = 6-note flag
  251 = 7-note flag
252-254 = Reserved
  255 = end of track
```

Byte 2:

Bit 0:	Tie forward flag
Bits 1-7:	Length of note in divisions (0—128)

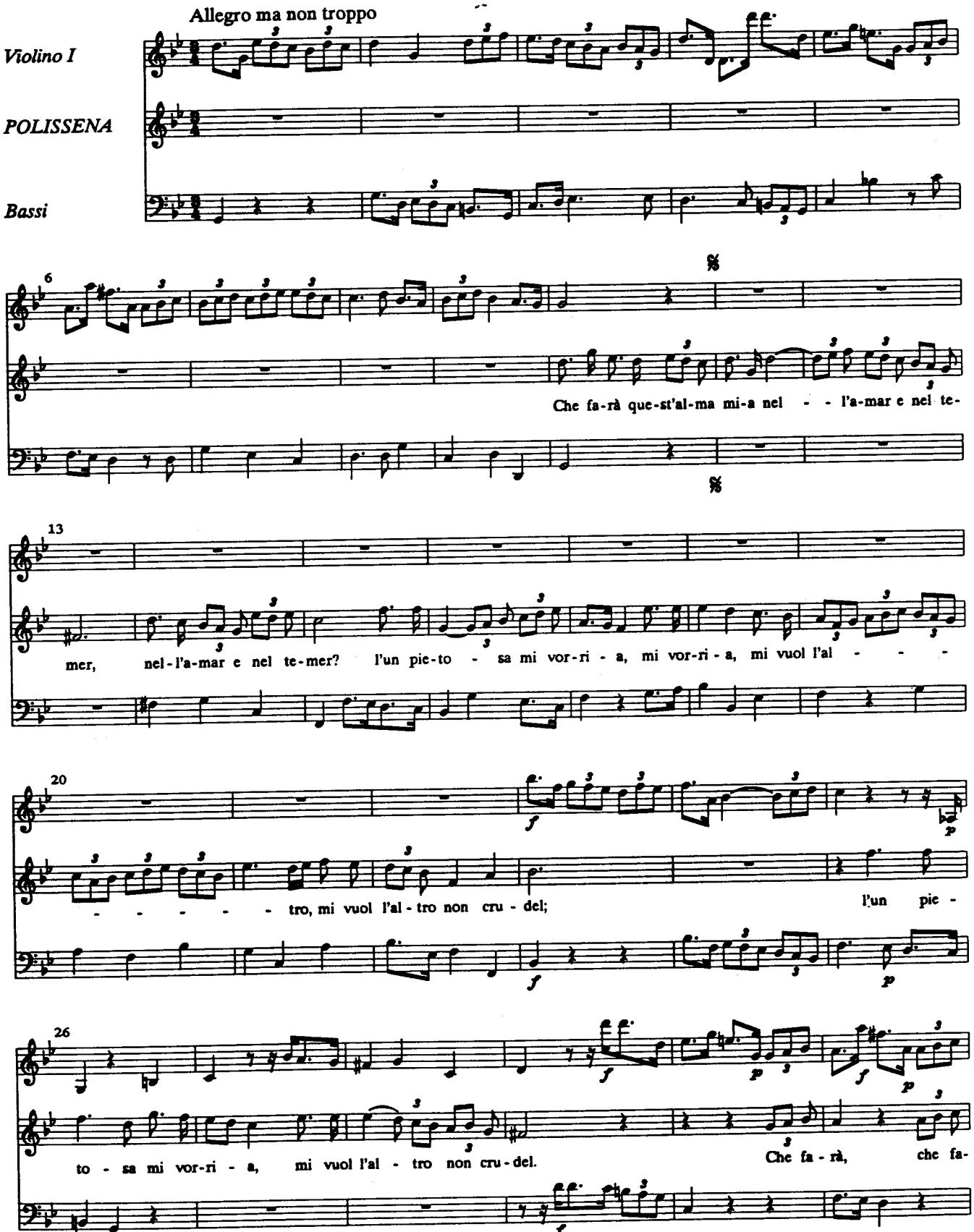
When there are two or more notes, the pitch values follow in successive bytes. Two pitches take up one word; three pitches take up two words; four pitches take up two words, etc.

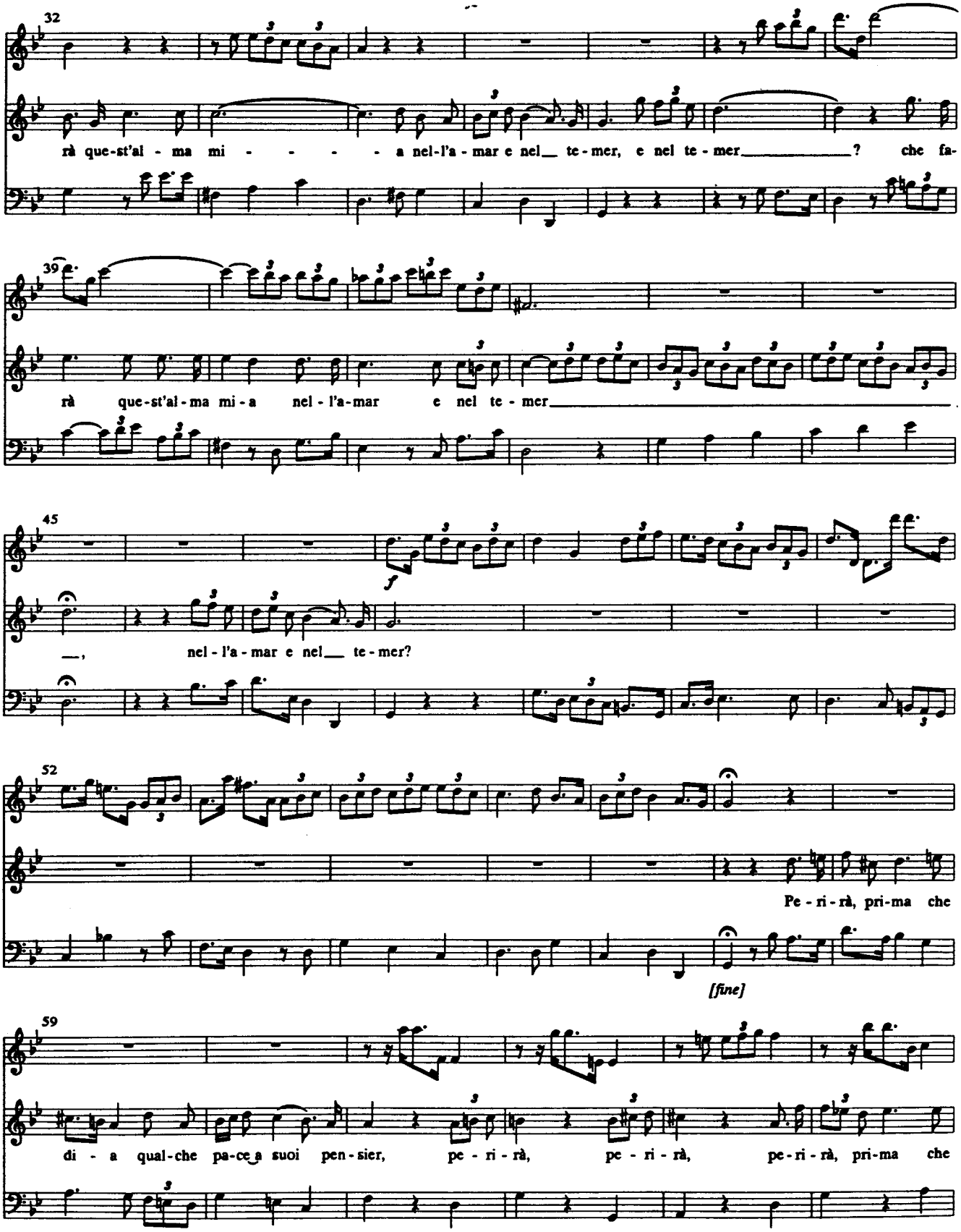
As successive data records fill up, new data records are created. The last data record in the movement (section) is padded with zeroes, so that the next movement (section) can start on a new record boundary.

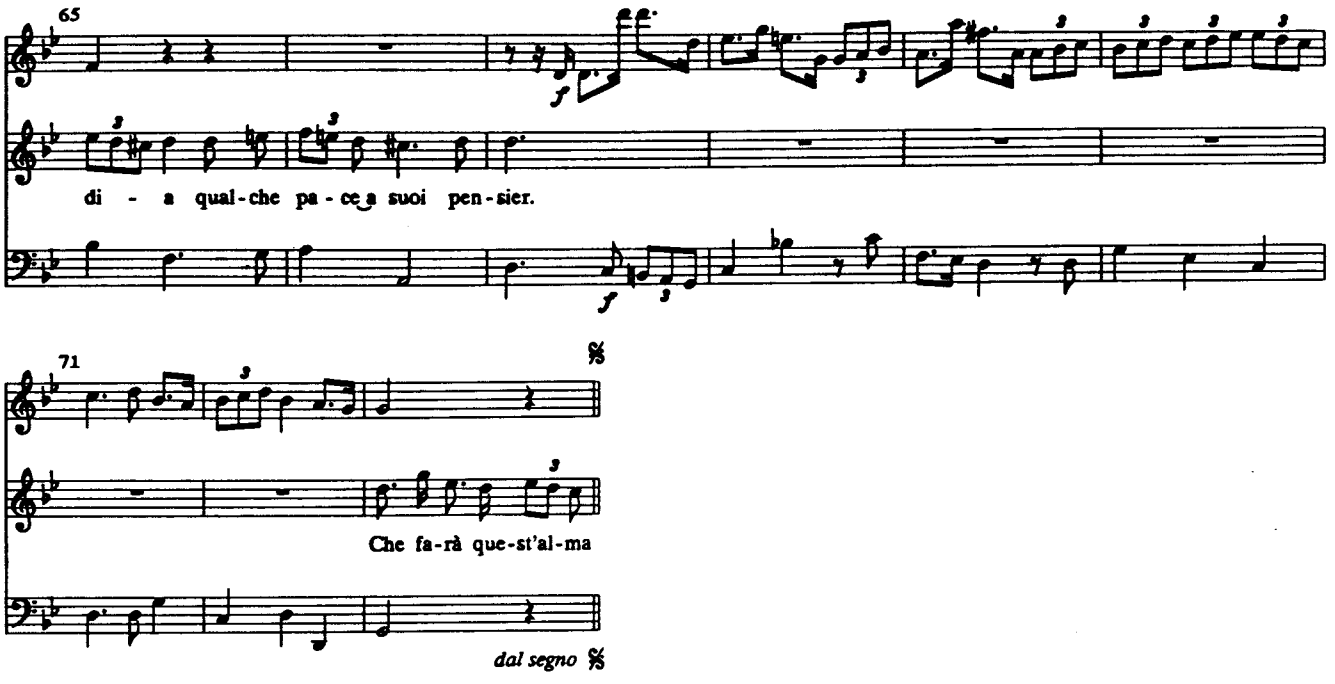
Graphic Information

(Stage 2)

Sample Score (Vocal): G. F. Handel, *Radamisto*, score for accompanied aria, bars 1-73




32
rà que-st'al - ma mi - - - - - a nel-l'a - mar e nel__ te - mer, e nel te - mer_____? che fa-
39
rà que-st'al-ma mi-a nel - l'a-mar e nel te - mer_____
45
f
—, nel - l'a - mar e nel__ te - mer?
52
Pe - ri - rà, pri-ma che
[fine]
59
di - a qual-che pa-ce a suoi pen-sier, pe - ri - rà, pe - ri - rà, pe-ri - rà, pri-ma che


65
f
di - a qual-che pa - ce a suoi pen-sier.
f
71
Che fa-rà que-st'al-ma
dal segno 𝄋

Input File for Vocal Part: G. F. Handel, *Radamisto*, Stage 2 file for vocal part

```
12/05/92 E. Correia
WK#:12b       MV#:48
Chrysander, 2nd Version December 1720 / Terence Best 1992..
Radamisto
Aria
POLISSENA
0 0 S
Group memberships: data
data: part 5 of 6
&
Initial conversion from stage 1 to stage 2
&
$  K:-2   Q:12   T:3/4   C:4   D:Allegro ma non troppo
rest  36
measure 2
rest  36
measure 3
rest  36
measure 4
rest  36
measure 5
rest  36
measure 6
rest  36
measure 7
rest  36
measure 8
rest  36
measure 9
rest  36
measure 10
D5      9        e.    d                        Che
G5      3        s     d                        fa-
Ef5     9        e.    d                        r\a8
D5      3        s     d                        que-
Ef5     4        e 3 d [     *                  st'al-
D5      4        e 3 d ]                        -
C5      4        e 3 d       !                  ma
measure 11       A
D5      9        e.    d                        mi-
G4      3        s     u                        a
D5     24-       h     d        -               nel-
measure 12
D5      4        e 3 d [     *                  -
Ef5     4        e 3 d ]                        -
F5      4        e 3 d       !                  l'a-
Ef5     4        e 3 d [     *                  mar_
D5      4        e 3 d ]                        _
C5      4        e 3 d       !                  e
Bf4     4        e 3 u [     *                  nel_
A4      4        e 3 u ]                        _
G4      4        e 3 u       !                  te-
measure 13
F#4    36        h.#   u                        mer,
measure 14
D5      9        e.    d                        nel-
C5      3        s     d                        l'a-
Bf4     4        e 3 u [     *                  mar_
A4      4        e 3 u ]                        _
G4      4        e 3 u       !                  e
Ef5     4        e 3 d [     *                  nel_
D5      4        e 3 d ]                        _
Ef5     4        e 3 d       !                  te-
measure 15
C5     24        h     d                        mer?
F5      9        e.    d                        l'un
F5      3        s     d                        pie-
measure 16
G4     12-       q     u        -               to-
G4      4        e 3 u [     *                  -
A4      4        e 3 u ]                        -
Bf4     4        e 3 u       !                  sa
C5      4        e 3 d [     *                  mi_
D5      4        e 3 d ]                        _
Ef5     4        e 3 d       !                  vor-
measure 17
A4      9        e.    u [                      ri-
G4      3        s     u ]\                     -
F4     12        q     u                        a,
Ef5     9        e.    d                        mi
Ef5     3        s     d                        vor-
measure 18
Ef5    12        q     d                        ri-
D5     12        q     d                        a,
C5      9        e.    d                        mi
Bf4     3        s     d                        vuol
measure 19
A4      4        e 3 u [     *                  l'al-
F4      4        e 3 u =                        -
G4      4        e 3 u ]     !                  -
A4      4        e 3 d [     *                  -
Bf4     4        e 3 d =                        -
C5      4        e 3 d ]     !                  -
Bf4     4        e 3 u [     *                  -
A4      4        e 3 u =                        -
G4      4        e 3 u ]     !                  -
measure 20
C5      4        e 3 d [     *                  -
A4      4        e 3 d =                        -
Bf4     4        e 3 d ]     !                  -
C5      4        e 3 d [     *                  -
D5      4        e 3 d =                        -
Ef5     4        e 3 d ]     !                  -
D5      4        e 3 d [     *                  -
C5      4        e 3 d =                        -
Bf4     4        e 3 d ]     !                  -
measure 21
Ef5    18        q.    d                        -
D5      3        s     d [[                     tro,_
Ef5     3        s     d ]]                     _
F5      6        e     d                        mi
Ef5     6        e     d                        vuol
measure 22
D5      4        e 3 d [     *                  l'al-
C5      4        e 3 d ]                        -
Bf4     4        e 3 d       !                  tro
F4     12        q     u                        non
A4     12        q     u                        cru-
measure 23
Bf4    36        h.    d                        del;
measure 24
rest  36
```

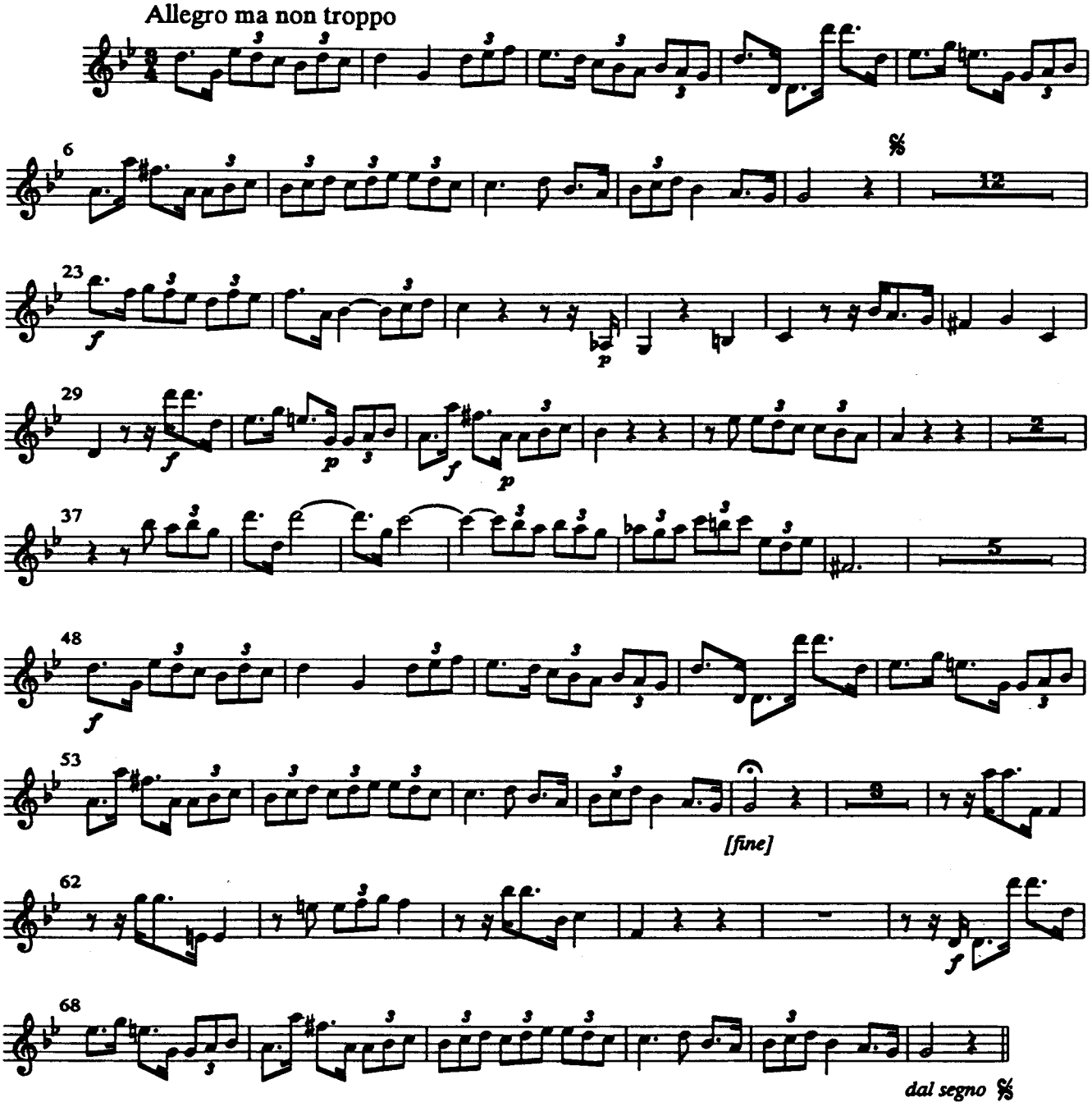
```
measure 25
rest  12        q                                       
F5    18        q.      d                       l'un    
F5     6        e       d                       pie-    
measure 26
F5    18        q.      d                       to-     
D5     6        e       d                       sa      
G5     9        e.      d                       mi      
F5     3        s       d                       vor-    
measure 27
Ef5    6        e       d       [               ri-     
D5     6        e       d       ]               -       
C5    12        q       d                       a,      
Ef5    9        e.      d                       mi      
Ef5    3        s       d                       vuol    
measure 28
Ef5   12        q       d               (       l'al-   
D5     4        e  3    d               )*      -       
C5     4        e  3    d       [               tro_    
Bf4    4        e  3    d       ]       !       _       
A4     4        e  3    u       [       *       non_    
Bf4    4        e  3    u       ]               _       
G4     4        e  3    u               !       cru-    
measure 29
F#4   24        h #     u                       del.    
rest  12        q                                       
measure 30
rest  12        q                                       
rest  12        q                                       
G4     4        e  3    u       [       *       Che_    
A4     4        e  3    u       ]               _       
Bf4    4        e  3    u               !       fa-     
measure 31
A4    12        q       u                       r\a8,   
rest  12        q                                       
A4     4        e  3    d       [       *       che_    
Bf4    4        e  3    d       ]               _       
C5     4        e  3    d               !       fa-     
measure 32
Bf4    9        e.      d                       r\a8    
G4     3        s       u                       que-    
C5    18        q.      d                       st'al-  
C5     6        e       d                       ma      
measure 33
C5    36-       h.      d               -       mi-     
measure 34
C5    18        q.      d                       -       
D5     6        e       d                       a       
Bf4    6        e       d                       nel-    
A4     6        e       u                       l'a-    
measure 35
Bf4    4        e  3    d       [       *       mar_    
C5     4        e  3    d       ]               _       
D5     4        e  3    d               !       e       
Bf4   12        q       d               (       nel_    
A4     9        e.      u               )       _       
G4     3        s       u                       te-     
```

```
measure 36
G4    18        q.      u                       mer,    
G5     6        e       d                       e       
F5     4        e  3    d       [       *       nel_    
G5     4        e  3    d       ]               _       
Ef5    4        e  3    d               !       te-     
measure 37
D5    36-       h.      d               -       mer?_   
measure 38
D5    12        q       d                       _       
rest  12        q                                       
G5     9        e.      d                       che     
F5     3        s       d                       fa-     
measure 39
Ef5   18        q.      d                       r\a8    
Ef5    6        e       d                       que-    
Ef5    9        e.      d                       st'al-  
Ef5    3        s       d                       ma      
measure 40
Ef5   12        q       d                       mi-     
D5    12        q       d                       a       
D5     9        e.      d                       nel-    
D5     3        s       d                       l'a-    
measure 41
C5    18        q.      d                       mar     
C5     6        e       d                       e       
C5     4        e  3    d       [       *       nel_    
B4     4        e n3    d       ]               _       
C5     4        e  3    d               !       te-     
measure 42
C5   12-        q       d               -       mer,_   
C5     4        e  3    d       [       *       _       
D5     4        e  3    d       =               _       
Ef5    4        e  3    d       ]       !       _       
D5     4        e  3    d       [       *       _       
Ef5    4        e  3    d       =               _       
C5     4        e  3    d       ]       !       _       
measure 43
Bf4    4        e  3    u       [       *       _       
A4     4        e  3    u       =               _       
G4     4        e  3    u       ]       !       _       
C5     4        e  3    d       [       *       _       
Bf4    4        e  3    d       =               _       
A4     4        e  3    d       ]       !       _       
D5     4        e  3    d       [       *       _       
C5     4        e  3    d       =               _       
Bf4    4        e  3    d       ]       !       _       
measure 44
Ef5    4        e  3    d       [       *       _       
D5     4        e  3    d       =               _       
Ef5    4        e  3    d       ]       !       _       
C5     4        e  3    d       [       *       _       
D5     4        e  3    d       =               _       
Bf4    4        e  3    d       ]       !       _       
A4     4        e  3    u       [       *       _       
Bf4    4        e  3    u       =               _       
G4     4        e  3    u       ]       !       _       
measure 45
D5    36        h.      d               F       _       
```

```
measure 46
rest  12        q
rest  12        q
G5     4        e  3  d  [     *          nel-
F5     4        e  3  d  ]                -
Ef5    4        e  3  d        !          l'a-
measure 47                     .
D5     4        e  3  d  [     *          mar_
Ef5    4        e  3  d  ]                _
C5     4        e  3  d        !          e
Bf4   12        q     d        (          nel_
A4     9        e.    u        )          _
G4     3        s     u                   te-
measure 48
G4    36        h.    u                   mer?
measure 49
rest  36
measure 50
rest  36
measure 51
rest  36
measure 52
rest  36
measure 53
rest  36
measure 54
rest  36
measure 55
rest  36
measure 56
rest  36
measure 57
*               C        [fine]
rest  12        q
rest  12        q
D5     9        e.    d                   Pe-
E5     3        s n   d                   ri-
measure 58
F5     6        e     d                   r\a8,
C#5    6        e #   d                   pri-
D5    18        q.    d                   ma
E5     6        e n   d                   che
measure 59
C#5    9        e.#   d  [                di-
B4     3        s n   d  ]\               -
A4    12        q     u                   a
D5     6        e     d                   qual-
A4     6        e     u                   che
```

```
measure 60
Bf4    3        s     d  [[               pa-
C5     3        s     d  ]]               -
D5     6        e     d                   ce\0+a
C5    12        q     d        (          suoi_
Bf4    9        e.    d        )          _
A4     3        s     u                   pen-
measure 61
A4    12        q     u                   sier,
rest  12        q
A4     4        e  3  d  [     *          pe-
B4     4        e n3  d  ]                -
C5     4        e  3  d        !          ri-
measure 62
B4    12        q n   d                   r\a8,
rest  12        q
B4     4        e  3  d  [     *          pe-
C#5    4        e #3  d  ]                -
D5     4        e  3  d        !          ri-
measure 63
C#5   12        q #   d                   r\a8,
rest  12        q
A4     9        e.    u                   pe-
F5     3        s     d                   ri-
measure 64
F5     4        e  3  d  [     *          r\a8,_
Ef5    4        e f3  d  ]     +          _
D5     4        e  3  d        !          pri-
Ef5   18        q.    d                   ma
Ef5    6        e     d                   che
measure 65
Ef5    4        e  3  d  [     *          di-
D5     4        e  3  d  =                -
C#5    4        e #3  d  ]     !          -
D5    12        q     d                   a
D5     6        e     d                   qual-
E5     6        e n   d                   che
measure 66
F5     4        e  3  d  [     *          pa-
E5     4        e n3  d  ]                -
D5     4        e  3  d        !          ce\0+a
C#5   18        q.#   d                   suoi
D5     6        e     d                   pen-
measure 67
D5    36        h.    d                   sier.
measure 68
rest  36
measure 69
rest  36
measure 70
rest  36
measure 71
rest  36
measure 72
rest  36
measure 73
D5     9        e.    d                   Che
G5     3        s     d                   fa-
Ef5    9        e.    d                   r\a8
D5     3        s     d                   que-
Ef5    4        e  3  d  [     *          st'al-
D5     4        e  3  d  ]                -
C5     4        e  3  d        !          ma
*               B        dal segno
mdouble         A
/END
```

Part Output: G. F. Handel, *Radamisto*, violin part from accompanied aria, Bars 1-73



Technical Description

Graphic (Stage 2) File Format

Introduction

The CCARH Musical Data Bases are full-text encodings. There are no tags in the data nor relational structures in the format. The basic aim is to encode raw rather than interpreted data. In the encoding process many situations requiring the resolution of ambiguity and contradiction arise, and in this sense the data is edited.

The development of the CCARH Data Bases is a long-term project which it is hoped will produce an electronic corpus of information that can serve the diverse needs of musicians and musical scholars for decades to come. As a non-profit research center, CCARH will make its data bases available by non-exclusive license.

Although parts of the format have been stable for several years, each new repertory seems to present a few unique features, and minor adjustments and accommodations are made from time to time. This description (November 1990) should therefore be regarded as provisional.

The following topics are covered:

Overview
- The Header
- The Main Section
- The Footnote Section
- Source Documentation

Data Types in the Main Section
- Musical Attributes
  - Field identifiers
- Musical Directions
  - Codes for musical direction data types
- Bar lines

Regular notes and rests
- Codes for note durations
- Codes for accidentals
- Codes for ties, slurs, and tuplets
- Codes for ornaments
- Codes for accidentals relating to ornaments
- Codes for technical indications
- Codes for articulations and accents
- Other indications and codes

Extra note in a chord
Grace notes, cue notes
- Codes for grace and cue notes

Extra grace/cue note in a chord
Figured harmony
- Codes for modifiers in basso continuo figuration

Forward and back space in time
Continuation line
Comments
End of music data or end of file
Summary of Control Codes

The Center welcomes comments on this source file format. Complete documentation will be made available with the first release of large data sets.

Walter B. Hewlett
November 1990

Overview

A CCARH source file is organized as a set of variable-length records. Each source file describes the music in one musical part of a musical work. A musical part may consist of one or more lines of music. For example, parts for Oboe 1 and Oboe 2 may be combined on one staff and therefore be considered as one musical part ("Oboes"). They may also be treated as two separate parts. Music on the grand staff may be considered as one or two parts. If musical notation or symbols cross between the staves of the grand staff, then the music on the grand staff must be treated as one musical part.

A source file has **three sections**. These are the header, the body of musical data, and an optional section for footnote data.

The Header

Header records 1-10 provide identifying information concerning the encoding process, the musical work, and the source(s) on which the encoding is based. These are currently assigned as follows:

Record 1: free
Record 2: free
Record 3: free
Record 4: [date] [encoder]
Record 5: WK#: [work number] MV#: [movement number]
Record 6: [source]
Record 7: [work title]
Record 8: [movement title]
Record 9: [name of part]
Record 10: free (we put in [mode], [movement type] and [voice])

Records 11 and 12 indicate relationships between this and other parts of the work. The format is

Record 11: Group memberships: [name1] [name2] . . .
Record 12: [name1]: part [n] of [number in group]
Record 13: [name2]: part [n] of [number in group]
. . .

The value of these records can be illustrated by the case in which there are files for *Oboe 1*, *Oboe 2*, and a combined *Oboes 1, 2*. The part called *Oboe 1, 2* would belong to a group called *Fullscore*, whereas the *Oboe 1* and *Oboe 2* parts by themselves would belong to the group called *Parts*. In addition, if the

oboe parts are musically independent from other instrumental parts, they might belong to a group called *Tracks*.

The names and arrangements of the groups and parts is left to the encoder. For each group to which a part belongs, there is a record (starting with Record 12) giving the group name and the sequential number of that part in that group. The arrangement of the parts into various groups allows flexibility in printing full scores, short scores, and individual instrumental parts. It makes possible the representation of parts for vocal soloists both as separate parts (for analysis) and in combination with parts for other soloists (*e.g.*, as in recitatives). It makes possible the representation of keyboard works both as single files (for printing) and as separate tracks (for analysis).

The Main Section

The main section of a source file follows directly after the header. The end of this main section is marked by a special end-of-movement record consisting of the five characters "/FINE". All records between the header and the /FINE record are data records (with the exception of records between comment designator flags). The order of these records is an integral part of the information contained in the file. The first character in each data record functions as a control code, describing the type of information contained in the record. For certain data types, such as notes, the first character also contains data information. This system of describing record types with a control code allows for significant future expansion of the source-file format. At the moment, there are 21 control codes and twelve data types.

The Footnote Section

The optional footnote section begins after the /FINE record. The end of the source file is marked by a special end-of-file record consisting of the four characters, "/END". If there is no footnote section, the /FINE record may be omitted. The format of the footnote section is left to the discretion of the encoder. The footnote section may contain musical data. In this case, a record with an exclamation mark (!) in Column 1 is used to signal that records in the musical data format will follow. A second record with an (!) in Column 1 will signal that the following records are no longer in the musical data format.

Source Documentation

Source-file data may come from several sources such as modern editions, early prints, or manuscripts. These sources may incorporate the amendments of multiple scribes, editors, and users. It is important that the source-file format be able to distinguish between these various sources and levels of information. Because of the complicated nature of this problem, several methods are provided for tagging data according to the source and/or layer of information within the source.

Group membership: As described earlier, each source file has a list of groups to which it belongs. In the case in which a particular work exists in several versions or has sources which differ substantially, these versions or sources may be represented in separate files, group memberships here being used to indicate which version(s) a track or part belongs to.

Level numbers: Within a particular source file, the principal technique for distinguishing between sources and/or levels of editing is the use of level numbers. Each data record has a column set aside for labelling the record as belonging to a particular level. In this manner, any musical attributes (*e.g.*, time signature, key, tempo designation), any musical directives (*e.g.*, dynamics, tempo changes, etc.), any musical notes or bar lines may be identified with a particular source or level of editing. In addition, a wide variety of musical notations which are attached to a note such as ornaments, articulations, slurs, phrase markings, fingerings, local dynamics, etc., may be assigned a level number. In this way, the same note may have editorial markings from several different levels attached to it. There are 35 possible levels, and the assignment of meaning to these levels is the responsibility of the encoder.

Footnotes: Every data record has a column set aside for specifying a footnote. The set of characters used in the footnote column and the assignment of their meaning is left to the discretion of the encoder. The purpose of the footnotes is to provide the encoder the opportunity to add additional information at various points in the data file. Such information might include alternate readings, additional measures of music, written out ornaments, or any discussion of the sources or the editorial process.

Group memberships are expressed in the course of creating particular kinds of scores. Footnotes can be appended in the usual manner to prints of the data. Level information can be conveyed in printed versions through the use of diverse fonts and font sizes.

The source-file *font designator* provides a method of specifying different fonts and thereby distinguishing the levels of information. In cases in which the font designator is introduced as part of a line of ASCII data (*e.g.*, "un poco f", with the "f" being in a different font from "un poco"), the character "!" is used, followed immediately by the font number (*e.g.*, un poco !4f, where 4 in this example might be the encoder's designator for the standard music font).

Data Types in the Main Section

For the encoding of musical information in the body of the file, twelve data types are currently in use. They are:

1. musical attributes (key, time signature, etc.)
2. musical directions (dynamics, etc.)
3. bar lines
4. regular notes and rests
5. extra note(s) in a regular chord
6. grace and cue notes
7. extra grace or cue notes in a chord
8. figured harmony
9. forward or back space in time
10. continuation line
11. comments
12. end-of-music (/FINE) or end-of-file (/END) marker

A control character in Column 1 indicates the data type of the record. The remainder of this document presents an item-by-item explanation of these types.

1. Musical Attributes **Control code = $**

Musical attributes include key signatures and time signatures. They are relevant to the correct interpretation of the note information. The elements of a musical attribute record are as follows:

Column 1: "$"
Column 2: level number (optional)
Column 3: footnote column (optional)
Columns 4-80: attribute fields

An attribute record may contain one or more fields. Each field is initiated by a *field identifier* and terminated by one or more blanks. In the case of clefs and directives, the field identifier may contain a number, which is the staff (1 or 2 currently) to which the clef or directive belongs. The absence of a number indicates staff number one.

Field identifiers

Field Identifier	Field Type	Data Type
K:	Key	Integer
Q:	Divisions per quarter note	Positive integer
T:	Time designation	Two integers
C:	Clef	Integer
X:	Transposing part	Integer
S:	Number of staves for part	Integer (default = 1)
I:	Number of instruments represented	Integer (default = 1)
D:	Directive (last field in record)	ASCII string

This is an example of an attribute record with multiple fields:

$ K:-2 Q:8 T:3/8 C:4 C2:22 S:2 D:Allegro ma non troppo

A. Key Codes (K:)

The numbers -7 to +7 are reserved for standard key signatures. Minus numbers are for flats; positive numbers are for sharps. The designation "-4" could refer to Ab Major or F Minor. The designation "0" could refer to C Major or A Minor. Other integer codes may be assigned at the discretion of the encoder.

B. Divisions per quarter note (Q:)

The duration of each pitch and rest is given as a certain number of time units (divisions). The relationship between time units and musical notation is specified by the number of time units in a quarter note. This parameter can be specified only at the beginning of a piece or directly after a controlling bar line. Specifying it elsewhere makes it difficult to combine separate parts for printing or for MIDI output.

C. Time designation (T:)

The time designation is given by two non-negative integers separated by a slash (/). The first integer is normally the time numerator and the second integer is normally the time denominator. Some common designations would be T:3/4 and T:12/8. Some special codes in current use are these:

1/1 = common time (C)
0/0 = alla breve
2/0 = simple 2
3/0 = simple 3

Other codes with a 0 denominator may be devised by the encoder to represent special time notations.

D. Clef codes (C:)

The clef code is a positive integer between 1 and 85. The tens digit specifies the clef sign, while the ones digit specifies the staff line to which the clef sign refers. The clef-sign codes are these:

0 = G-clef
1 = C-clef
2 = F-clef
3 = G-clef transposed down (modern tenor clef)
4 = C-clef transposed down
5 = F-clef transposed down
6 = G-clef transposed up
7 = C-clef transposed up
8 = F-clef transposed up

The "0" for the G-clef is a default and need not be stated.

The representation of staff line numbers is:

1 = highest line
5 = lowest line

E. Transposing part (X:)

This integer (+ or -) indicates a transposing interval, if there is one, and/or a doubling of the part an octave lower. The base-40 system is used. 23 means the music sounds a fifth higher than it is written; -23 means the music sounds a fifth lower. Adding 1000 to the number indicates a doubling of the part an octave lower (*e.g.*, vc and bass on the same part, or 8' and 16' sound on an organ pedal line.)

F. Number of staves (S:)

This integer (1 or 2 currently) indicates the number of staves the part or parts will be written on. This number can change within a movement. The number of staves will automatically be set to 2 if a "C2:" or a "D2:" is encountered.

G. Number of instruments represented (I:)

This integer (1 or more) indicates the number of independent instruments represented by the part. If this number is more than one, certain printing conventions will hold:

(1) notes with the same stem direction will be combined into one chord.

(2) if more than one voice is represented in a measure on a staff, then each voice will follow its own set of accidentals within the measure.

H. Directive (D:)

This ASCII string data group is terminated by the end of the record. For this reason, if a record contains a directive, the directive must be the last field of the record. Directives may contain font designators.

2. Musical Directions **Control code = ***

Musical directions consist of verbal and graphic information related to performance and may include rehearsal numbers, pedal indications, crescendos and diminuendos, dynamics expressed by letters, octave transpositions, and so forth. The column allocations and data types are as follows:

Column 1: "*"

Columns 2-5: blank

Columns 6-8: optional forward offset (measured in units of duration and right justified in the field). Use of this field allows the encoder to place a musical direction at a division that does not otherwise contain a musical record.

Columns 9-12: blank

Columns 13-15: footnote and level information

- Column 13: footnote flag (blank = none)
- Column 14: level number (optional)
- Column 15: track number (necessary if there are two or more wedges, sets of dashes, 8ve-up transpositions, etc.)

Column 16: blank

Columns 17-18: type of direction (one or two letters)

Codes for musical direction data types

Data Type	Code	Meaning
Rehearsal numbers	A	Rehearsal numbers/letters (ASCII string)
Verbal directions	B	Right-justified ASCII string
	C	Centered ASCII string
	D	Left-justified ASCII string
Wedges	E	Begin wedge
	F	End wedge
Letter dynamics	G	Letter dynamics (ASCII string)
Dashes	H	Begin dashes (after words)
	J	End dashes
Pedal (pianoforte)	P	Begin pedal [Ped.]
	Q	Release pedal [*]
Octave shifts	U	Shift notes up (*e.g.*, by 8ve)
	V	Shift notes down (*e.g.*, by 8ve)
	W	Stop shift

Notes:

B, C, and D may be combined with E, F, H, and J

E and F may be combined with B, C, D, and G

G may be combined with E, F, H, and J

H and J may be combined with B, C, D, and G

Column 19: location flag (optional)

" " = indication below staff
"+" = indication above staff (may be used by types A, B, C, D, E, F, G, H)

Column 20: blank

Columns 21-23: numerical parameter (optional)

Types A, B, C, D, G: optional font designator
Types E, F: wedge spread
Types U, V: shift size when not 8ve

Wedge spread is measured in tenths of staff line space. 10 units = space between two staff lines.

Column 24: staff number (" " = 1)

Used in the case of music represented on more than one staff.

Columns 25—: ASCII word string (used in A, B, C, D, G)

Some examples of musical directions are:

1. *cresc.* - - - - - - - - ff

Starting record:	DH	*cresc.*	
Ending record:	JG	ff	

2. f <decreasing wedge> p

Starting record:	GE	15	f
Ending record:	FG	0	p

3. <increasing wedge> p

Starting record:	E	0	
Ending record:	FG	15	p

3. Bar lines **control code = m**

Column 1:	"m"	
Columns 2-7:	"easure"	= regular bar line
	"dotted"	= dotted bar line
	"double"	= (light) double bar line
	"heavy1"	= heavy bar line
	"heavy2"	= light-heavy double bar
	"heavy3"	= heavy-light double bar
	"heavy4"	= heavy-heavy double bar
Column 8:	blank	
Columns 9-12:	optional bar number for this bar (left-justified)	
Columns 13-15:	footnote and level information	
Column 13:	footnote flag (blank = none)	
Column 14:	level number (optional)	
Column 15:	blank	
Column 16:	blank	
Columns 17-80:	flags	

Flag codes are read from left to right. The character "&", followed by a digit (1..9, A..Z), is used to indicate a specified data level. All codes to the left of the first "&" belong to the initial data level. These codes are used:

*	= non-controlling bar line
~	= continue ~ ~ ~ across bar line
A	= segno sign at bar
start-end#	= start ending #
stop-end#	= stop ending #
disc-end#	= discontinue ending # line
:|	= repeat backward
|:	= repeat forward

Codes are separated by one or more blanks.

Bar lines are divided into two types—controlling and non-controlling. *Controlling* bar lines run throughout an entire score. In this respect, they mark the beginning of a new global measure. *Non-controlling* bar lines may not serve for line breaks or page breaks. The designation of a bar line as non-controlling is to some extent left to the discretion of the encoder. For example, a double bar in the middle of a normal measure could be considered controlling or non-controlling. However, in a case such as that of the minuet in Mozart's *Don Giovanni*, where the score uses three different meters simultaneously, the non-aligned bar lines must be designated as non-controlling.

4. Regular notes (A..G) and rests (r) **Control code = A, B, C, D, E, F, G, or r**

Columns 1-4: pitch (Cff0 to B##9, where C4 = Middle C) or rest

Column 5: blank

Columns 6-8: duration (right-justified)

Column 9: tie flag

" " = no tie
"-" = tie from this note to the next

From the pitch and duration data in columns 1 to 9, it is possible to reproduce aural output of the musical part.

Columns 10-12: blank

Columns 13-15: footnote and level information, track number

Column 13: footnote flag (blank = none)
Column 14: level number (optional)
Column 15: track number (optional)

Where more than one musical line is represented in a part (*e.g.*, Oboe I and II, or keyboard music), it is essential for purposes of analysis to know for each note (or chord) the musical line or "track" to which the note belongs. In some cases this is "interpretive" information, provided as a service by the encoder.

Column 16: blank

Columns 17-22: note description

Column 17: note duration type (table on next page)

Codes for note durations

Code	Name
L	longa
b	breve
w	whole
h	half
q	quarter
e	eighth
s	sixteenth
t	thirty-second
x	sixty-fourth
y	128th
z	256th
d	dot independent of note
v	void (no note in this position)

Column 18: dot flag

" " = no dot
"." = single dot
":" = double dot

Column 19: actual accidental flag ("actual accidental" is the accidental as it appears in the musical source)

Codes for accidentals

Code	**Meaning**
#	sharp [♯]
n	natural [♮]
f	flat [♭]
x	double sharp [x]
X	sharp-sharp [♯♯]
S	natural-sharp [♮♯]
F	natural-flat [♮♭]
&	flat-flat [♭♭]

Columns 20-22: designation of time modification

This requires two digits, separated by a colon (:). For standard cases, such as triplets (3:2), the colon and the second digit are usually omitted. The numbers 10-35 are represented by the letters A-Z.

Column 23: stem direction

d = down
u = up
" " = no stem

Column 24: staff number (" " = 1)

Used in the case of music represented on more than one staff.

Column 25: blank

Columns 26-31: beams (up to six levels = 256th note)

[= start beam
= = continue beam
] = end beam
/ = forward hook
\ = backward hook

Column 26 is for eighth beams
Column 27 is for sixteenth beams
. . .

Columns 32-43: other notations

Codes are read from left to right. The character "&", followed by a digit (1..9, A..Z), is used to indicate a specified data level. All codes to the left of the first "&" belong to the base data level.

The following codes are somewhat arbitrary. They have been chosen for representing common musical notation for Western music from the sixteenth through the nineteenth centuries. The encoding scheme is not complete but may be augmented and/or altered to meet the special requirements of the music being encoded.

Codes for ties, slurs, and tuplets

Code	Meaning
-	tie
(	open slur1
)	close slur1
[	open slur2
]	close slur2
{	open slur3
}	close slur3
z	open slur4
x	close slur4
*	start tuplet
!	stop tuplet

Codes for ornaments

Code	Meaning
t	tr.
r	turn
k	delayed turn
w	shake
~	trill extension [~ ~ ~]
c	continue wavy line
M	mordent
j	slide (Schleifer)

Codes for accidentals relating to ornaments
(must follow directly after ornament)

Code	Meaning
s	sharp (ss = double sharp)
h	natural
b	flat (bb = double flat)
u	next accidental is below ornament

Codes for technical indications

Code	Meaning
v	up bow
n	down bow
o	harmonic
0 (zero)	open string
Q	thumb position (cello)
1,2,3,4,5	fingering (pianoforte)

Codes for articulations and accents

Code	Meaning
A	vertical accent [/\]
V	vertical accent [\/]
>	horizontal accent
.	staccato
_	tenuto or marcato
=	line with dot under it
i	spiccato
,	breath mark

Other indications and codes

Code	Meaning
S	arpeggiate (chords)
F	upright fermata
E	inverted fermata
p	piano (also pp, ppp)
f	forte (also ff, fff)
m	mezzo (mp, mf)
Z	sfz
Zp	sfp
R	rfz
^	editorial accidental
+	cautionary accidental
:	next fingering is a substitution, *e.g.*, 5:4 = 5-4

Columns 44-80: multiple lines of text (the verses of vocal music) set off by |

Example: Deck|See|Fast (from "Deck the Halls")

5. Extra note in a chord **Control code = " "**

Column 1: blank
Columns 2-5: pitch (see regular note)
Columns 6-8: blank
Column 9: tie flag
Columns 10-42: same as for regular note

6. Grace notes, cue notes **Control code = g or c**

Column 1: g = grace note
c = cue note (grace-note size, but written in time)
Columns 2-5: pitch/rest (see regular note)
Columns 6-7: blank
Column 8: note type

Codes for grace and cue notes

Code	Note type
0	eighth note with slash
1	256th note
2	128th note
3	64th note
4	32nd note
5	16th note
6	eighth note
7	quarter note
8	half note
9	whole note
A	breve

Column 9: blank
Columns 10-80: same as regular notes

7. Extra grace/cue note in a chord **Control code = g or c**

Column 1:	g or c (same as 6.1 above)
Column 2:	blank
Columns 3-6:	pitch (see regular note)
Column 7:	blank
Column 8:	note type (same as 6.8 above)
Column 9:	blank
Columns 10-80:	same as other grace and cue notes

8. Figured harmony **Control code = f**

Column 1:	f = figured harmony
Column 2:	number of figure fields
Columns 3-5:	blanks
Columns 6-8:	advance of figure division pointer

Figures take their position from the first regular note that follows the figure records. In the case in which the figures change during the duration of a note, the advancing parameter (columns 6-8) is used to indicate the elapsed time between changes. In the case in which a figure appears after a note has sounded, the "blank" figure is used as a place holder to advance the figure division pointer.

Columns 9-12:	blank
Columns 13-15:	footnote and level information
Column 13:	footnote flag (blank = none)
Column 14:	level number (optional)
Column 15:	blank
Column 16:	blank
Columns 17—:	figure fields

The figure fields are set off by one or more blanks. Figure numbers may extend from 1 to 19. They may be modified in front by #, n, f, and x. They may be modified afterward by #, n, f, +, \, and x. The #, n, f, and x signs may stand alone as figures. A "b" indicates a blank figure. This is used as a place holder in a list and also to start a continuation line with no figure. The first figure field is for the top of the figure list.

In the figure list #4 $\begin{matrix}6\\2\end{matrix}$, the 6 would be represented in the first field.

The codes for signs and modifiers are shown below.

Codes for modifiers in basso continuo figuration

Sign	Meaning
#	sharp
n	natural
f	flat
x	double sharp
+	augment (used with figures 2, 4, and 5)
\	augment (used with figures 6 and 7)
-	short line in
_	long line from previous figure
b	blank (place holder)

9. Forward and back space in time **Control code = i, b**

Columns 1-5: "irest" = forward space (invisible rest)
"back" = back space
Columns 6-8: duration to skip forward or back up
Columns 9-12: blank
Columns 13-15: footnote and level information
Column 13: footnote flag (blank = none)
Column 14: level number (optional)
Column 15: blank
Column 16: blank
Column 17: pass number (optional)

This feature can be used to express parallel action in the same part (*e.g.*, keyboard music).

10. Continuation line **Control code = a**

Column 1:	a = append to previous line
Columns 2-16:	blank
Columns 17-80:	continuation of previous line

11. Comments **Control codes = & and @**

Column 1: @ = this line is a comment
& = enter comment mode. All subsequent records are comments until another record with a "&" in Column 1 appears. "&" in Column 1 acts like a toggle switch between "data" and "comment" mode.

12. End of music data or end of file **Control code = /**

Column 1:	/
Columns 2-5:	"FINE" = end of music data
	"END" = end of file

In the case in which there is no footnote section, the "/FINE" record may be omitted.

Summary of Control Codes

Code	Meaning
" "	extra note in chord
$	controlling musical attributes
&	comment mode toggle switch
*	musical directions
A	regular note
B	regular note
C	regular note
D	regular note
E	regular note
F	regular note
G	regular note
/	end of music data or end of file
@	single line comment
a	append to previous line
b	backspace in time
c	cue-size note
f	figured harmony
g	grace note
i	invisible rest
m	bar line
r	regular rest