DOCUMENT RESUME

ED 041 031 24 TE 499 852

AUTHOR Lincoln, Harry B.
TITLE Development of Computerized Techniques in Music Research with Emphasis on the Thematic Index. Final Report.
INSTITUTION State Univ. of New York, Binghamton.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO BR-8-B-089
PUB DATE Jul 69
GRANT OEG-0-8-080089-4581
NOTE 15p.

EDRS PRICE EDRS Price MF-$0.25 HC-$0.85
DESCRIPTORS *Computer Graphics, *Computer Programs, *Data Bases, Data Processing, Electronic Data Processing, *Indexes (Locaters), Information Retrieval, Information Storage, *Music, Programing, Research Methodology

ABSTRACT

This project comprises the second year's work of a 3-year research plan to develop computer programs, necessary data, and procedures for using the computer in a special field of musicology known as thematic indexing, and to explore the use of these techniques in other areas of music. During the second year, the research centered on three areas: (1) continuation of encoding of incipits (opening theme of a piece) of 16th-century materials (This data bank, now containing more than 40,000 themes, serves as testing material for validating programs as well as providing a source of information for other musicologists.) (2) continued development of special programs for information retrieval, pattern matching, and organization of data for publication by direct photo copy of computer output, (3) development of special type faces of music symbols for use on the high speed computer printer. Preliminary tests with the new type were successful. (Appendices include examples of two stages of computer printing of music notation.) See also ED 027 609. (Author/MF)

Final Report

Project No. 8-B-089

Grant No. OEG 0-8-080089-4581

DEVELOPMENT OF COMPUTERIZED TECHNIQUES IN MUSIC RESEARCH WITH EMPHASIS ON THE THEMATIC INDEX

Harry B. Lincoln
Department of Music
State University of New York at Binghamton
Binghamton, N.Y. 13901

July 1969

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Acknowledgments

The research described in this report could not have been carried out without the cooperation and professional help of various persons. Mr. Robert Roberson, Director of the Computer Center at State University of New York at Binghamton made the services of the Center available for the project and encouraged it in many ways. When Mrs. Cay Gill, who did much of the early programming the project, left the Center for another position she was replaced by Mr. Christian Granger, who is both a capable programmer and a trained music theorist. His combination of skills and enthusiasm for the project has been a major asset in its development.

Table of Contents

SUMMARY

This project comprises the second year's work of a three-year research proposal. (The first year's work was carried out under USOE Project No. 7-8276, Grant No. 1-7-078272-3898.) The purpose of this research is to develop computer programs, necessary data, and procedures for using the computer in a special field of musicology known as thematic indexing, and to explore the use of these techniques in other areas of music. The project was not concerned with the use of the computer in music composition.

In the first year of the project it was clearly established that the computer can be an effective tool in expanding the usefulness of the thematic index, itself an established tool in musicological research for over fifty years. When done by hand the thematic index entails an enormous amount of effort in the preparation and sorting of thousands of cards. Past efforts have been restricted to small repertories--automation makes possible the exploration and systematic searching of many thousands of themes.

During the past year the research has centered on three areas:

1. Continuation of encoding of incipits (opening theme of a piece) of sixteenth-century materials. This data bank of themes serves as testing material for validating programs as well as providing a source of information for other musicologists. At the conclusion of the second year of the project the data bank had more than 40,000 themes.

2. Continued development of special programs for information retrieval, pattern matching, and organization of data for publication by direct photo copy of computer output.

3. Development of special type faces of music symbols for use on the high speed computer printer. The work was carried out by the Glendale Laboratory of the IBM Corporation. The special type characters were completed and installed near the end of the research period. Preliminary tests with the new type were successful. The coming year will require the development of special programs to effectively format data and print musical themes in a manner satisfactory for photo reproduction and publication.

The investigator reported on this project in a paper to the International Federation of Information Processing Societes, meeting at Edinburgh, Scotland, in August 1968. A lecture describing this project was given to the Conference on Computer Applications to the

Humantiies and Social Sciences at Wooster Polytechnical Institute on November 15, 1968. A series of three lectures on this work were given at the University of Maryland in February, 1969, and a paper describing some prospects for the research as it pertains to musicology was given to the District of Columbia Chapter of the American Musicological Society on February 18, 1969. In the Spring of 1969 work was completed on editing a volume Computers and Music, a collection of papers by twenty-two contributors, to be published by Cornell University Press in late 1969.

BACKGROUND FOR THIS RESEARCH PROJECT

This project comprises the second year's work of a three-year research proposal. The purpose of this research is to develop computer programs, necessary data, and procedures for using the computer in a special field of musicology known as thematic indexing, and to explore the use of these techniques in other areas of music. The project was not concerned with the use of the computer in music composition.

The background work for this research project was summarized in the final report of the first year of work, funded under Project No. 7-8276, USOE Grant No. OEG 1-7-078272-3898. For the benefit of the reader who does not have access to that report, the following background information is quoted from that report:

"The thematic index, in its many forms, has long been a common tool for musicological research. It is a listing of the opening themes of the complete works of a particular composer, or, more rarely, of a particular genre, and usually consists of the first seven to twelve notes, known as the incipit. It serves several functions to the researcher, functions analagous to those of concordances and indices of first lines of poetry in the field of literature. It also provides data for the analysis of musical style or changes in style of a composer or genre.

"As developed by traditional methods, the thematic index is limited in its uses and applications. Traditional or "hand" methods of development include the preparation of elaborate card files, lists, catalogs, etc. With small repertories these can be manageable, but with larger bodies of material they become completely unwieldly. Publication of the results of the research calls for setting of type and autographing the music, both expensive processes.

"A small number of musicologists have realized that the whole system would lend itself to automation. The necessary research to develop automated procedures calls for time, effort, ideas, and financial support. The present research project is the most ambitious by far of any attempted in this area in America or European musicology.

"Prior to the present research, the investigator developed a small pilot project to test the feasibility of using the computer for indexing and analysis. The sixteenth-century Italian vocal repertory known as the frottole served as material for this study. In carrying out the pilot study, attempts were first made to encode music for the computer by simply entering the letter names of the notes of a melody. It became apparent that while this would provide data quickly for very

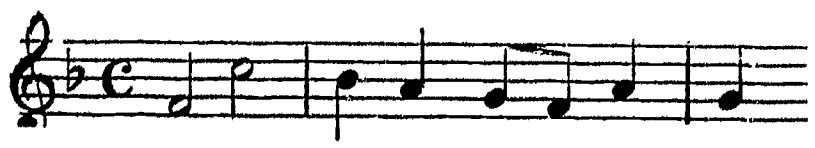
limited computer analysis, it did not take rhythm into consideration and thus would be of little value for other studies. During the pilot project stage it was decided to use a music representation known as the Ford-Columbia Representation for encoding melodies. This representation, developed by Mr. Stefan Bauer-Mengelburg, President of Mannes College, was intended for a project in music graphics. Although this work in the printing of music remains unfinished, the music representation has become accepted as an excellent tool for music analysis and indexing by several researchers. The Ford-Columbia representation is discussed more fully in the description of the present research below.

The following steps were carried out in the pilot project:

1. The 4000 incipits of the frottole repertory (from 1000 four-voice pieces) were encoded and keypunched in the Ford-Columbia Representation. Each piece was identified by a six-digit serial number and the investigator's initials. The use of initials will permit quick identification and separation of repertories contributed by other researchers.

2. One thousand cards containing basic information (composer, title, etc.) for each piece were keypunched.

3. A computer program was written which extracted the interval sequence from the Ford-Columbia Representation. Example 1 illustrates the generation of this information. Below the example the Ford-Columbia Representation is shown and on the following line the computer-generated set of numbers describing the interval sequence. Repeated notes are not considered for the important reason that a composer may, for example, borrow a melody and may change a whole note to two half notes in order to accomodate a new text. Our interest is in the over-all contour of the melody.



```
!G !K1- !MC 2H  6 / 5Q  4  (3 2) 4Q / 3Q
              +5  -2  -2  -2 -2 +3  - 2
```

4. The interval sequences were arranged in numerical order. Groups of identical sequences were single spaced in the printout. A double space separated the identical groups. This permits quick identification of matching contours. The data from the basic information cards, stored on computer disk, was matched by serial number and printed with the interval sequence.

5. In each group of matching contours a check of the Ford-Columbia Representation or the original music provided confirmation of genuine matches.

"The pilot project showed that heretofore unidentified duplications and borrowings in the frottole repertory do indeed exist and proved the feasability of applying these techniques to much larger repertories. The investigator reported on the pilot study in a paper to the American Musicological Society's annual meeting at New Orleans in December 1966."

During the first year of the funded project several procedures were carried out. A large amount of data was acquired, endcoded, and keypunched. Programs were written to permit greater varieties of information and to permit more effective formats in the print-outs. Other musicologists were invited to submit incipits to the data bank. By the end of the first year the total number of incipits in the data bank exceeded 25,000. During this period a Conference on Thematic Indexing was held in Washington, D. C., on November 4, 1967. Twenty-three musicologists and music librarians, all with an interest in thematic indexing, were in attendance.

RESEARCH PROCEDURES OF THE 1968-1969 PROJECT

Following the pattern established in the first year of the project, several procedures were carried out simultaneously. Briefly there were:

1. Continued acquisition, encoding and keypunching of data.

2. Development of necessary hardware, programs and procedures to permit printing of music on the high-speed computer printer.

3. Continued merging of materials from other researchers. The total data bank to date numbers about 40,000 cards of information.

4. Continued comparison of print-outs and pattern matchings with original music to determine effectiveness of the research design.

These four areas are considered now in detail:

1. Continued acquisition, encoding, and keypunching of data.

Emphasis has continued on sixteenth-century Italian materials found in modern editions and in microfilms of sixteenth-century editions of music obtained from European libraries. The university library has funded the purchase of microfilms (about $500) used during the past year. The microfilms revert to the library collection when the encoding is completed. A log is kept of serial number assigned to each piece of music. Appendix A illustrates a summary page of this log, showing groups of serial numbers assigned to collections. A detailed log of every single assignment is also kept. The summary log is published to provide information in response to inquiries about the data bank.

2. Development of necessary hardware, programs and procedures to permit printing of music on the high-speed computer printer.

This work has represented a major effort during the past year. As indicated in the "Conclusions and Recommendations" section of the first year's report, it became clear that the effectiveness of any large-scale thematic indexing by computer would be limited by the difficulty in rapidly reading the alphanumeric output. A means of printing the actual music notation would greatly facilitate interpretation of data as well as make possible economical publication of results.

Consultations were held with the typography department of the Glendale Laboratories of the IBM Corporation at Endicott, N. Y. They designed a set of characters for the print train which will permit printing open and closed note heads (on a line or on a space), stems, flags, and some of the rests used in music. This work was financed ($2400) by the Computer Center at the university. During the months that this work was under way the Center's academic programmer, Christian Granger, developed the necessary programs for printing of the staff notation. This proved to be rather complex since the computer paper moves in only one direction and thus the program must first develop a two-dimensional array in storage and then properly distribute the printing of the characters as the paper moves through. The program has also been devised so that, for example, there is more space following a whole note than following a half note, etc. This spacing of materials is important to the appearance and legibility of music.

Appendices B and C show two stages in the development of the music printing programs as worked out by Mr. Christian Granger in the Computer Center. In Appendix B the logic of the program has

been worked out using conventional symbols. In Appendix C the new type characters are used for the first time. In the next phases of the work, to be developed during the summer of 1969, special pre-printed staff paper will be printed (under a contract with Moore Business Forms).

3. Continued mergings of materials from other researchers.

There have been a number of inquiries from other musicologists wishing to merge their materials with the data bank. This has proved especially attractive to PhD thesis candidates, of which two (Lynn Trowbridge of Illinois and Mrs. J. Courtney Adams of Maryland) have submitted data. Professor Johannes Riedel (Minnesota) is preparing data on Ecuadorian sources. In return for providing material for the data bank these persons will benefit from citation of borrowing and duplications and other pertinent bibliographical material as comparisons are made between the submitted material and that already held in the data bank.

4. Continued comparison of print-outs and pattern matchings with original music to determine effectiveness of the research design.

Final confirmation of a borrowing or duplication in a repertory depends on a comparison of the original music of the two pieces in question. From our experience in the pilot project we know that we are going to find interesting examples of borrowing within and between repertories. We have delayed lengthy examination of thousands of apparent duplications pending the development of the special music printing arrangement. It is also expected that the two-dimensional arrays programmed for the printing of the music will provide data for treating these arrays as wave forms of the music itself, that is as a picture of both pitch and rhythm, and that it will be possible to arrange these in meaningful orders. This will be a major undertaking in the coming months of this research.

RESULTS OF THE PROJECT

1. A very large data bank of early music (primarily sixteenth-century Italian vocal music sources) has been developed and is available to other researchers as well as serving as test material for computer programs written under this project. About 40,000 themes are stored in the data bank along with composer and title information on over 8,000 compositions. Listings by composer, title and other formats are available to researchers.

2. Effective programs have been developed and are operative to compute interval sequences from incipits, to arrange and format data according to interval sequence and/or composer, title, etc. These programs were summarized in the final report of the first year's project.

3. A music typography for the high-speed computer printer has been successfully developed. The special type characters are now available from the IBM Corporation to all researchers. The necessary programs for translating data in the Ford-Columbia Representation into correct formats of the type characters to print the music have been written. These programs are available from the Computer Center at SUNY/Binghamton, New York.

It should be noted that because this report covers the second year of a three year project, each of the above phases is on-going and work on them will continue next year.

CONCLUSIONS AND RECOMMENDATIONS

1. It has become increasingly evident that the worth of this project lies in two areas --- the acquisition of data and the development of new techniques. An effort has been made to develop enough data to make it possible to thoroughly test the usefulness and validity of the programs (new techniques) developed. The acquisition of data is an ongoing process to which other researchers can contribute in the future since the size of a data bank is limited only by storage capacity of computers and the amount of music which has been written.

Although the acquisition of data has continued throughout the project the emphasis has shifted during the second year toward the development of new techniques, including not only programs but also hardware development. This new music type characters now available from IBM as a result of this project should increase the interest and capabilities of other music researchers.

2. The new type characters make possible the printing of 3x5 cards (library catalog size) containing music notation as well as textual information. It is recommended that music librarians interest themselves in the possibilities of this new printing technology. There are many possible applications, including folk song repertories, popular music, lieder, etc. to name but a few. The investigator has reported on this work to the New York State Chapter of the Music Library Association.

3. The most important programming application on which work must yet be done is in the area of precise pattern matching, that is a

matching of melodies which considers rhythm as well as pitch. The programs of the two-dimension arrays used in printing music with the new type characters will in themselves provide a basis for the necessary work in pattern matching. It is expected that this will be a major effort in the third year of the project.

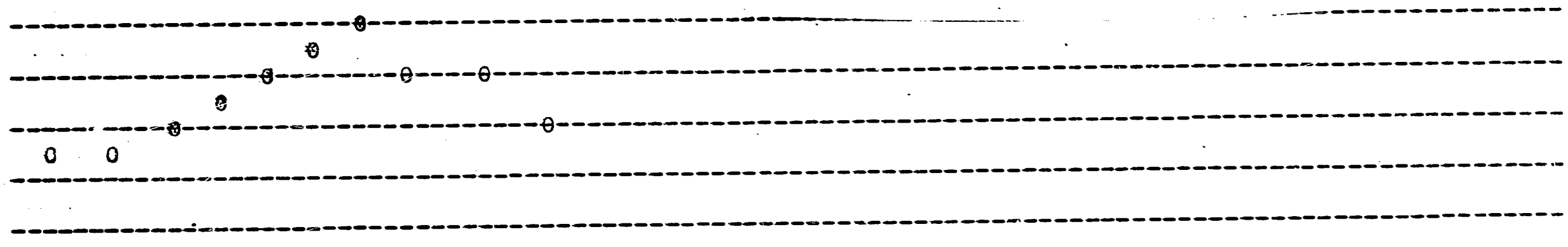
4. Finally, it is recommended that efforts be made to utilize the computer typewriter terminal as a means of entering data and, possibly, for performing certain calculations. The State University at Binghamton was the first in the country to use the new programming language APL, recently released by IBM. Binghamton is joining with the University of Alberta, through SHARE, to develop an APL users group. APL is a powerful, flexible language, designed for mathematics and science. It also has great possibilities in certain areas of music theory and analysis. Mr. Granger, computer center programmer, has written interesting and sophisticated programs for calculating combinatorial possibilities in 12-tone rows. Although this work appears to be outside the province of the present project it is one which should be explored for possible applications. The present project is tied to the entering of data by cards. Card input is being replaced by other devices in many computer applications and the possibilities for changes in music research need to be explored.

APPENDIX A

Sample page of log.

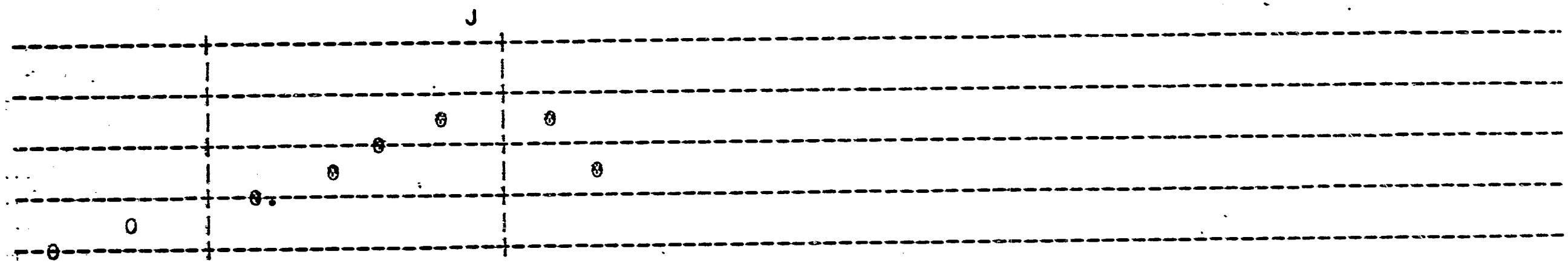
Composer (abbrev.)	Serial #	Library Call No.	Source (abbreviated)
GAFFURIO	9801-9985	M2/A538/V1 +	ARCHIV MUS METRO MED
GAFFURIO	9927-9937	M2/A538/V4	ARCHIV MUS METRO MED
GAFFURIO	9938-9985	M2/A538/V5	ARCHIV MUS METRO MED
G PALESTRINA	10001-12882	M3/P15/V1 +	CASIMIRI ED V1-30
	15501-525		RISM 1536/7
	15526-553		RISM 1537/7
	15554-581		RISM 1537/9 (1533/2)
	15582-603		RISM 1586/9
	16001-056		RISM 1541/14
	16057-124		RISM 1543/19
	16127-175		RISM 1543
	16176-16224		RISM 1502/2
	17601-628		RISM 1547/14
	17901-923		RISM 1547/21
A AGRICOLA	18001-245	M3/A35/V1 +	CMM 22 OPERA OMNIA
CLEMENS	19501-20093	M3/C6/V5-8/9-11	CMM 4 OPERA OMNIA
LUPI	19561-62	M3/C6/V5-8	CMM 4 WITH CLEMENS
SERMISY	19563	M3/C6/V5-8	CMM 4 WITH CLEMENS
MANCHICOURT	19564-565	M3/C6/V5-8	CMM 4 WITH CLEMENS
A WILLAERT	19566	M3/C6/V5-8	CMM 4 WITH CLEMENS
LUPI	19837	M3/C6/V5-8	CMM 4 WITH CLEMENS
LUPI	19845	M3/C6/V5-8	CMM 4 WITH CLEMENS
CLAUDRIN	19852	M3/C6/V5-8	CMM 4 WITH CLEMENS
SANDRIN	19852	M3/C6/V5-8	CMM 4 WITH CLEMENS
	20401-431		RISM 1550/19
ANON	20501-516	M2/A538/V7	ARCHIV MUS METRO MED
ANON	20517-538	M2/A538/V9	ARCHIV MUS METRO MED
H ISAAC	20539-565	M2/A538/V10	ARCHIV MUS METRO MED
VAN WERBEKE	20567-598	M2/A538/V11	ARCHIV MUS METRO MED

First stage in development of printing of music notation.

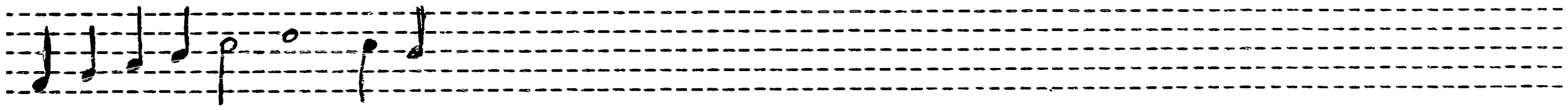


```
1H 2 / 3Q. 4E 5Q 6J / (6 4) 7Q.
21H    22H    |      23Q.   24E    25Q    26JQ   |      26E    24E    $$$$$$$
```

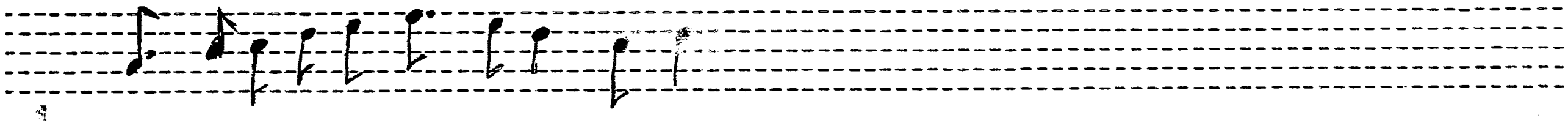
COMPOSER: GPALEST
TITLE: QUODCUMQUE VINCLIS
INTERVAL SEQ: 2+02+02+02+02-03+04



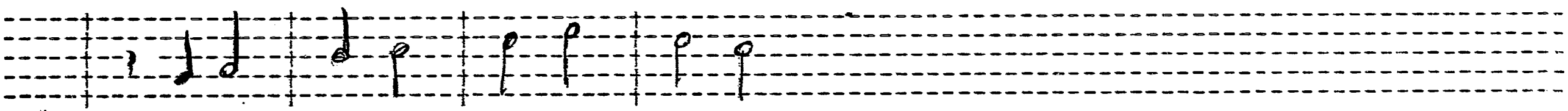
Later stage in development of printing of music notation. This shows the first use of special music type characters on a computer printer.



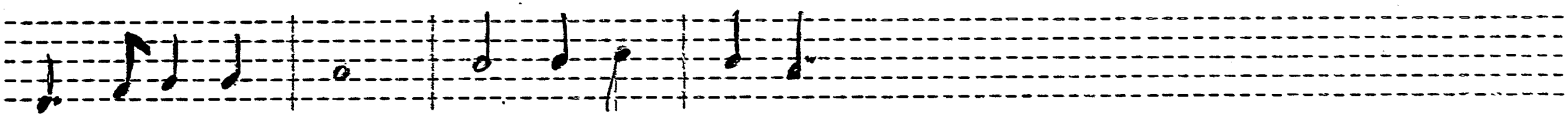
COMPOSER: TITLE: INTERVAL SEQ: +02+02+02+02+02-02-02



COMPOSER: GPALEST TITLE: UT, RE, MI, FA, SOL, LA INTERVAL SEQ: +02+02+02+02+02-02-02



COMPOSER: GPALEST TITLE: DE BEATE INTERVAL SEQ: +02+02+02+02+02-02-02



COMPOSER: GPALEST TITLE: PATER NOSTER INTERVAL SEQ: +02+02+02+02+02-02-02