OPTICAL CHARACTER-RECOGNITION OF PRINTED MUSIC: A REVIEW OF TWO DISSERTATIONS

AUTOMATIC RECOGNITION OF SHEET MUSIC. By Dennis Howard Pruslin. Sc. D. Dissertation, Massachusetts Institute of Technology, 1966.

COMPUTER PATTERN RECOGNITION OF STANDARD ENGRAVED MUSIC NOTATION. By David Stewart Prerau. Ph. D. Dissertation, Massachusetts Institute of Technology, 1970.

Readers of *Perspectives* scarcely need be reminded of the pre-eminence of the written-musical domain (i. e., that domain of musical experience in which music is presented visually in one or another system of musical notation) in musicology: before Edison composers could not produce records of their work in the sounded-musical domain, and other domains of musical experience such as the tactile domain utilized in the Braille system have been employed comparatively infrequently; and even after Edison various extra-musical considerations (such as copyright law and the relatively high cost of sound-processing machinery) have joined with tradition to keep the written-musical domain a principal mode of non-transient musical communication. Within this domain various systems of musical notation have achieved various degrees of currency at various places and times, but of all these systems one—the current common musical notation ('CCMN' for short)—has dominated: virtually all music printed has been printed in one or another 'dialect' of CCMN: even music originally noted in another system generally has been transcribed into CCMN before printing.

In recent years digital computers have become more efficient and more prevalent, so that today, at least in computationally well-developed parts of the world, it no longer is unreasonable to delegate, or to plan to delegate, musical processes to electronic computing machinery. Of course, many musical processes do not involve previously recorded musical compositions: perhaps it is to the comparatively early success of a few such computer-mediated processes that an unfortunate synecdochic misidentification of 'computer music' with 'synthesizing sound through the use of a digital computer' has arisen.[1] But (and of this too readers will be well informed) central to musicology are processes that do involve prior musical compositions, and for the full delegation of these processes to computing machinery the relevant compositions must be put into computer-acceptable form. Human key-punchers can transcribe from CCMN onto (say) punch cards (at Princeton University the Masses of Josquin were so transcribed, at a rate of approximately 20 minutes per printed CCMN page), but as this task clearly requires no intelligence beyond that with which machines can be endued it is only natural to consider

[1] See, for instance, p. 477 of the 1971-72 calendar of the University of Calgary.

constructing an artifact that can scan printed CCMN pages, recognize and discriminate amongst the 'primitive symbols' of CCMN there arrayed, and finally notify the result of this recognition in computer-acceptable form.[2] Although optical character-recognition (OCR) machines that 'read' printed text (in Roman and Cyrillic alphabets) have been operational for several years, the dissertations under review report—so far as I am aware—the first and only substantial endeavours to apply OCR technology to the 'reading' of printed CCMN. Both authors have essayed to solve less than the entire problem, so the work of each should be judged by its extensibility to, rather than by its non-realization of, an actual working machine.

Dr. Pruslin chose to limit attention to a subsystem of CCMN in which music is noted in just one measure on only two parallel staffs. Chords sharing a common stem are allowed on each staff to consist of up to four notes, of which no two adjacent notes are more than two staff-positions apart; but 'the sometimes occurring case in which two parallel melody lines exist for the right hand, left hand, or both, is not handled' (p. 18). The subsystem lacks rests, accidentals, half-notes, whole-notes, and notes attached to more than one stem; eighth-notes and notes of lesser note-duration are admitted only if beamed to an adjacent note. Clefs, time-signatures, grace-notes, dynamic marks, phrase marks, and other 'special signs' of CCMN are disallowed.

Pruslin's recognition scheme begins with a scanning device that converts a measure of music (noted in this subsystem) to a facsimile array of about 65,000 black-or-white points that are stored straightforwardly in a digital computer. A computer program processes this array to sharpen the focus both of horizontal lines (which may come to be recognized as staff-lines or beams) and of vertical lines (which may come to be recognized as stems or bar-lines); when in proper focus the staffs are located and the widths of the staff lines are computed.

Pruslin then isolates 'contour traces' by subtracting electronically both the staff-lines and the vertical lines and segmenting the residue so that—roughly speaking—within certain size limitations each contour trace encloses a minimal area that contains adjacent black points but is bordered entirely by white points. Each contour trace is potentially either a 'note cluster', i. e., a note or a dialectally permitted chord of two to four notes, or one or more beams used to connect eighth-notes or notes of lesser note-duration. A large portion of the dissertation is devoted to detailed presentation of the particular algorithms used to classify contour traces.

After recognition of the note clusters and the beams, a comparatively simple procedure assigns note-durational values (e. g., 1/4) to the note clusters. Pruslin does not go on to compute the pitches that the notation indicates, but

[2]Some requirements that such a device will have to fulfill are stated in my 'An essay toward specification of a music-reading machine' (Princeton, 1963, mimeographed; reprinted 1971 by the American Musicological Society (Greater New York Chapter) on pp. 151ff. of their ill-edited volume *Musicology and the Computer* (City University of New York Press)).

from his results (obtained using the TX-0 and IBM 7094 computers) the staff-positions of notes are readily determinable, and from this stage other algorithms can take over.[3]

Pruslin's reports of his experiments are encouraging. 'The note cluster recognition procedures were tested using a total of 153 note clusters. With the critical value of template match set at 70%, only one error, the rejection of a note cluster occurred' (pp. 88-89). Similar success is reported for other recognition procedures constituting Pruslin's scheme. However, one must balance this encouragement against the small size of the experimental sample (which appears to have been selected from just one musical publication) and the absence of evidence that Pruslin's recognition scheme provides a sensible basis for the ultimately wanted machine that recognizes all of CCMN.

Indeed, Dr. Prerau at first planned merely to extend Pruslin's methods to cover the parts of CCMN that Pruslin had excluded.[4] But upon finding 'that Pruslin's preprocessing erased or distorted most music symbols other than quarter-notes and beamed notegroups, the objects that his program was designed to recognize' (p. 42), Prerau chose a different approach. After scanning and digitizing two to three measures of a musical score, his computer program locates 'fragments'—contiguous black regions that are present either between a pair of adjacent staff-lines or just above or just below a staff. Next these 'fragments' are 'assembled', according to rules that connect certain adjacent fragments, into 'components' that—roughly—correspond either to isolated music symbols (e. g., an accidental or a note not beamed to any adjacent note) or to a group of beamed notes. The recognition of components as one or more particular music symbols is a multi-stage process, involving size ('a natural is a little taller than a flat' (p. 120)) as well as syntax ('a sharp can appear . . . as an accidental, to the left of a note . . . [or] in a key signature, to the right of a clef or a bar-line' (p. 144)).

Prerau limits attention to a subsystem of CCMN in which each of only two parallel staffs bears monolynear[5] music composed of notes, rests, treble and bass clefs, certain time-signatures, 'sofons'[6], and dots of prolongation—but not tempo indications, dynamic or phrase marks, or certain other 'special signs'. This CCMN dialect is instanced by many measures of an Edwards reprint of a Breitkopf & Härtel publication of Mozart's *Twelve Duets for Two Wind Instruments*, K. 487, which has served as the sole source for Prerau's ex-

[3]Some of these algorithms are presented in my 'A system for the automatic reduction of musical scores', *Papers Presented at the Seminar in Mathematical Linguistics, vol. 6 (1960),* on deposit at Widener Library, Harvard University.

[4]Both dissertations were supervised at least in part by Professor Murray Eden.

[5]I call a written-musical composition *monolynear* if its sounded-musical equivalent is monophonic. In monolynear music there are no chords or hiatuses, so that in general the sum of the note-durations of each note and rest in a measure of monolynear music is equal to the value of the time-signature affecting that measure.

[6]Prerau's acronym for 'sharp or flat or natural' (but not the double-sharp, which has been excluded from the dialect, and possibly not the double-flat or a sharp or flat preceded immediately by a natural).

periments. The recognition rules that Prerau adopts (for instance, to distinguish amongst 'sofons': 'The part of the component containing the top point is the rightmost 2/3 for a sharp, the leftmost 1/3 for a flat or natural; the part of the component containing the bottom point is the leftmost 1/2 for a flat, the rightmost 1/2 for a natural' (p. 210)) thus are derived from measurements of the same music Prerau's program proceeds to recognize; nonetheless, the reported experimental result that all dialectally permitted symbols in twenty or so measures from the Edwards reprint of K. 487 were recognized correctly is an impressive achievement. His program, called DO-RE-MI, is written for the IBM 7094 computer, and produces output in the Ford-Columbia representation of CCMN,[7] from which, at least potentially, another program can compute the represented pitches. DO-RE-MI is said to be constructed so that insertion of further algorithms to recognize portions of CCMN not included in Prerau's subsystem can be accomplished with relative ease: however, these algorithms have yet to be conceived.

The question 'How simple is it to extend a computer program that recognizes music issued by a certain firm at a certain time so that the program also recognizes music engraved (or printed) by other firms or at other times?' is closely related to the question 'How many 'founts' of music-engravers' punches have been widely employed?': unfortunately, neither question has yet been answered by any systematic investigation. But I understand, from a valuable conversation with Mr. H. Edmund Poole who is much concerned with certain historical aspects of music printing, that much German, U. K., and U. S. music engraving of the late 19th and early 20th centuries utilized punches and other tools produced by Messrs. C. G. Röder of Leipzig. Any resultant standardization of manufacturing style can be only of benefit to the designer of the ultimate OCR system for recognizing all printed CCMN, who will have to plan not only for a variety of different 'founts' but also for music symbols of differing size, for absence of expected spacing between symbols, for varying qualities of ink and paper, etc. The usually cited literature on music printing—e. g., William Gamble, *Music Engraving and Printing* (London, Pitman, 1923) or Karl Hader, *Aus der Werkstatt eines Notenstechers* (Vienna, Waldheim-Eberle, 1948)—will provide him only with little help.[8]

Perhaps the greatest accomplishment of the authors is that, as a result of their work, the logic of a machine that 'reads' multiple parallel staffs bearing polylynear[9] printed music in at least one 'fount' and size can be seen to be no further than another couple of M. I. T. dissertations away. Quite possibly such dissertations may get completed before much thought is directed toward deciding what wisely to do with the masses of musical data that an operational

[7]See Stefan Bauer-Mengelberg, 'The Ford-Columbia input language', printed on pp. 48ff. of the American Musicological Society publication mentioned in footnote 2.

[8]In general Pruslin and Prerau appear merely to have skimmed the literature on musical topics. Whilst this has not retarded their progress, careful readers will find passages in both dissertations where for want of musical knowledge full interdisciplinarity was not achieved.

[9]The written-musical analogue of 'polyphonic'.

OCR system could make available for computer processing. It is remarkable, nonetheless, that (of all things) this technology may cause return of musicologists' attention to the core concepts of their field which constitute musical theory.

Murray Eden has included a summary of Dr. Pruslin's work in an article.[10] Dr. Prerau presented salient parts of his work in a report to the 1971 Fall Joint Computer Conference.[11] This report was judged the 'best paper' submitted and—by way of recognition—he was awarded a plaque.

—Michael Kassler

[10] Murray Eden, 'Other pattern-recognition problems and some generalizations' in Paul A. Kolers and Murray Eden (eds.), *Recognizing Patterns* (Cambridge, Mass., M. I. T. Press, © 1968), pp. 196ff.

[11] David S. Prerau, 'Computer pattern recognition of printed music', *Proc. Fall Joint Computer Conference, vol. 39 (1971)* (Montvale, N. J., AFIPS Press), pp. 153ff.