

Pattern Recognition DATA BASES

1.1.1 Machine Imprinted Alphanumeric Characters – Dr. H. F. Ryan, Calspan Corp./U.S. Postal Service

An alphanumeric character data base (normalized version) of 100,000 samples of 66 character classes. (Also known as the CAL-U.S. Postal Service Alphanumeric Character Data Base.) Thresholded binary images of segmented centered mixed-font, machine-imprinted characters. Resolution is 24 x 24. Magnetic tape, 9 track, 2 reels, 1600 BPI.

Price: \$123.75 (\$75.50 with furnished tapes)
Member's discount price: \$99 (\$60 with furnished tapes)

1.1.1A Machine Imprinted Alphanumeric Characters – Dr. H. F. Ryan, Calspan Corp./U.S. Postal Service

An unnormalized version of an alphanumeric character data base which contains approximately 50,000 characters constituting a training data base (CALTRN) and 50,000 characters constituting a testing data base (CALTST). Magnetic tape, 9 track, 2 reels, 800 BPI.

Price: \$123.75 (\$75.50 with furnished tapes)
Member's discount price: \$99 (\$60 with furnished tapes)

1.1.2 Machine Imprinted Multilevel Characters – James S. Gibbons, Electronic Sciences Division, U.S. Postal Service.

An alphanumeric character data base of 32,000 (16 level) multi-level characters, extracted from 2100 test mail pieces each in a 24 x 24 array covering an area of 0.144 x 0.144 square inches. Complete information as to upper/lower case, font style, percent reflectance, color print and background accompanies each character on magnetic tape. Suitable for machine independent research as well as OCR testing. Magnetic tape, 4 reels test characters, 1 reel set up characters, 9 track, 800 BPI.

Price: \$262.50 (\$150.00 with furnished tapes)
Member's discount price: \$210.00 (\$120.00 with furnished tapes)

1.2.1 Handprinted Numeric Characters – Dr. A. L. Knoll, Honeywell Information Systems, Data Systems Division

The data base consists of 50 samples of each numeric character generated by 9 different authors. Simple printing rules were specified but not always followed. The samples were selected from those contributed. The images are binary with a resolution of 25 x 21. Punched cards.

Price: \$41.25
Member's discount price: \$33

In order to encourage research in the field of pattern recognition, the IEEE Computer Society's Technical Committee on Machine Pattern Analysis has begun collecting data bases from a variety of sources. These data bases, including substantial back-up documentation, may be ordered by using the form at the back of the issue.

Discounts off the data base list prices are available to IEEE members and members of the American Federation of Information Processing Societies' constituent societies.

When ordering, you may elect to send us your own blank tapes; if you do, be sure they are in good condition and have no other data recorded on them.

If you have a data base that you wish to contribute to the Technical Committee on Machine Pattern Analysis, please contact Dr. J. B. McFerran, Sperry-Univac, P.O. Box 3525, St. Paul, Minnesota 55165.

1.2.2 Handprinted FORTRAN Alphanumeric Characters – Dr. John H. Munson, Stanford Research Institute

The data base consists of two parts, with each part on a reel. The first part contains 3 alphabets of 46 characters, corresponding to the non-blank character set of the basic FORTRAN language, hand-printed by each of 49 authors making a total of 3 x 46 x 49 = 6,762 patterns.

The second part has 2,999 characters printed by a single author. There are 920 characters made up of 20 alphabets of 46 characters each; the remaining 2,079 characters are taken from fragments of actual coding sheets. The images are binary with a 24 x 24 resolution. Magnetic tape, 7 track, 2 reels, 556 BPI.

Price: \$116.75 (\$75.75 with furnished tapes)
Member's discount price: \$93 (\$60 with furnished tapes)

1.2.3 Handprinted Alphanumeric Characters – Dr. W. H. Highleyman, Sombers Associates.

There are approximately 2300 samples of alphanumeric characters. The images are binary with a resolution of 12 x 12. Punched cards.

Price: \$55
Member's discount price: \$44

1.2.4 Handprinted Numeric Characters – Hiroshi Genchi, Tokyo Shibaura Electric Co., Ltd./Toshiba Research and Development Center

The data base consists of 10,000 hand written numeric characters collected from live as well as experimental mail throughout Japan. The data base is contained on two



To order: Use the multipurpose order form at the back of the issue.

magnetic tapes with each tape having 5,000 characters. Images are binary with a resolution of 36 x 50. A single pattern consists of 56 words. Magnetic tape, 7 track, 2 reels, 800 BPI, 6 bits per character.

Price: \$115.25 (\$68.75 with furnished tapes)
Member's discount price: \$93 (\$55 with furnished tapes)

1.3.1 Cursive Script – Dr. L. D. Harmon, Bell Telephone Laboratories

The data consists of 52 cursive script sentences. The resolution for each sentence is 256 (vertically) x 2048 (horizontally). The images are binary. Magnetic tape, 7 track, 1 reel, 200 BPI.

Price: \$60.50 (\$42.25 with furnished tapes)
Member's discount price: \$50 (\$33 with furnished tape)

1.3.2 Cursive Script – Dr. R. W. Ehrich, ECE Department, University of Massachusetts

Handwritten 7-letter words recorded sequentially. A 400 word dictionary has been written by 3 writers and the first 64 words have been written by 4 additional writers. Each word consists of up to 256 data points. The data is on 7-channel magnetic tape, odd parity BCD, 3200 characters per record, 800 BPI.

Price: \$60.50 (\$42.25 with furnished tapes)
Member's discount price: \$50 (\$33 with furnished tape).

3.1.1 Speech Data Base – Dr. Harris Drucker, Monmouth College, John W. Preusse, Fort Monmouth

1090 examples of fourteen classes of phonemes produced by twenty speakers. Two-track, 1/4", magnetic audio tape with five minutes of speech and a 1 kHz timing signal; also, a paper tape identifying locations of phonemes on the audio tape. Recording speed 15 ips.

Price: \$40
Member's Discount Price: \$30