



PC TAP

Consumer Report

Desktop Scanners

ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

Report #5
January 1990

EPA
220/
1990.16



PC Technology Assessment Program
EPA National Data Processing Division
Information Centers Branch - RIC II, MD-35
Research Triangle Park, NC 27711
Telephone: (919) 541-0568 (FTS) 629-0568



IRM TAP
EPAX 3-1-1
9-1-1 1993
3000

ELB - 61950

PC TAP CONSUMER REPORTS

From the Editor's Desk

Our study of desktop scanners is highlighted in this *PC TAP Consumer Report*. Like our project to look at ways to perform graphics file transfers, this study grew along the way. Originally the objective was to assess just the DEST scanner that was available on the EPA PC contract. As you will see in the following pages, before the project was completed information was included about thirteen scanners. We also looked at a number of scanning front-end options, including both software-only and hardware/software-combination products for both the IBM PC and Apple Macintosh environments.

One of the reasons why our study kept expanding is the tremendous amount of attention scanning is currently getting in the industry. It's not so much that scanning technology is new, but the really good systems have been so expensive that they were out of reach for the typical desktop application. In the past couple of years, however, user demand for good scanning equipment has intensified the competition to provide such a capability while also bringing prices down to a more realistic range. Sounds like a familiar scenario, doesn't it? In any case, throughout the study more products kept coming to our attention, and we felt obliged to investigate as many as possible.

A second reason for the growth of our scanner study was that we at PC TAP just got so *into* it! Scanning is a fascinating topic, and the more we dug into it the more it grabbed us. Also, as we talked with other folks about scanning we learned about more users who have scanners, and everyone was anxious to have their device represented in our report.

We think our scanning study grew for good reasons, and that our report is better for the increased information it contains and the greater number of products it covers as a result of that growth. Certainly the input provided by the various participants resulted in a more comprehensive report than would otherwise have been possible. It has been an interesting report to research and write, and we hope you enjoy it. Due to the length of the scanner report, *Open Forum* does not appear in this issue.

David A. Taylor
PC TAP Coordinator

DESKTOP SCANNING

Introduction

Although this study was interesting for the PC TAP staff, it has also been somewhat frustrating. The frustration comes from the fact that scanning technology is getting so much attention in the industry and is changing so fast that it's hard to keep up. The more you learn about the process and about available products, the more you suspect there is that you haven't uncovered yet. New products keep cropping up everywhere, and at least one of those we're reporting on has announced an upgrade to the version we tested. But we're discovering that this is all part of the "technology assessment business"—playing catch-up with the industry.

Scanning: What's It All About?

As happens when you dig into most aspects of technology application, your vocabulary must be enhanced before you can explore the world of desktop scanning. It doesn't take long to find out that *scanning* is what a scanner does. And a scanner is a device that scans. Sound like technospeak doubletalk? It really isn't, it's just that the scanning process itself isn't all that complicated. A camera provides a good analogy: you use a camera to take a picture. Everyone can understand that process and what results from it. Well, a scanner "takes a picture" too. But what happens after the scanner captures the picture can get involved.

You point a camera and snap the shutter to capture photographic images of your choice on a roll of film. And when you've exposed an entire roll of film, you take it to a photo processor to have it developed. The result is a group of photographs. Scanners capture images too, but the camera analogy breaks down immediately after the capture takes place. That's because the scanner is a *computer peripheral*, while the camera is a stand-alone device. So rather than immediately recording the image (like the camera does on the film), the scanner simply passes it along to the host computer for further action. From that point on, the scanner is out of the loop and the material you've scanned is in your computer's memory waiting for you to do something with it.

We don't mean to infer, however, that capturing the image in the first place is insignificant. The wide range of capabilities and prices represented in the scanner marketplace gives some insight into the potential sophistication of these devices. While a basic desktop scanner (which may or may not be shipped with some front-end software) can be purchased for as little as \$1,000, a realistic cost estimate to equip yourself to scan text and graphics is roughly three times that, or about \$3,000, assuming you already have the computer to drive it all. One source lumps these document-scanning systems into the "low end" category that are widely used for desktop publishing and "typically sell for less than \$5,000" (*"Scanner Application Primer," Information Center, August 1989, p. 12*).

"Mid-range systems" generally are more powerful, more sophisticated versions of the low-end systems. They offer faster processing and heavier-duty equipment for a wider range of office applications, and can cost from \$5,000 to \$30,000. High-end scanners are designed for round-the-clock production use. Such systems can scan, enhance, compress, and capture images at a rate of about one per second, and they can accommodate a variety of physical document types. High-end systems cost approximately \$100,000. Then there's a "super high-end" category that we won't even go into that's in the \$250,000 ballpark.

If you're shopping for a system in the mid-, high-, or super-high-end category, don't waste your time reading further. This report is confined to the "low-end" category of scanning equipment. "Low-end" in this case doesn't mean *inferior*; it just signifies that the equipment in this group isn't as

sophisticated or as powerful as the more expensive gear in the higher categories. Low-end scanning equipment is well suited for office use and desktop publishing, where a very high percentage of scanning applications are found.

Text versus Images

We said earlier that the scanning device is out of the loop after the image has been captured. What then? Like so many things in the world today, it depends; and what it depends upon is the type of material you're processing. In the world of desktop scanning, you scan one of two things: *text* or *images*. We'll get into all the nuances of each of these processes later, but in general terms it all really boils down to whether your dealing with *words* or *pictures*. (Of course, no techie worth his or her salt would ever stoop to using such mundane terms.)

Word Processing

Let's talk about the processing of words (or, more properly, *scanning text*) first. This is a much more complex application than is apparent at first glance. In scanning parlance, the process of transforming a page of typed or printed text into a machine-readable form is called *optical character recognition* (OCR). Obviously, software is required to perform this process, and scanners often--but not always--are sold *without* such software. So, in addition to the cost of a scanner, you might have to buy an OCR package if you want to scan text. Basic OCR software is programmed to recognize certain character sets. The more capable a given package is in this regard, the more expensive it tends to be. In practice, the scanned page is held in memory while every single character is compared with those the software "knows" (this process is called *matrix matching*) to build a file containing ASCII text or, if your software has the capability, in the format of a word processing package. Matrix matching is suitable for recognizing text produced on typewriters, line printers, letter-quality printers, and (ostensibly) dot matrix printers.

A step up from matrix matching technology is required when you want to scan typeset material like books, magazines, and other professionally printed materials that usually contain a number of different type styles and sizes. Tackling the problem of character recognition in this environment requires more powerful software with more sophisticated capabilities. Using a process called *feature extraction*, which is based on the principle that each character has distinctive physical characteristics, such software packages examine the *features* of each scanned symbol and generate the appropriate character. Sometimes this is referred to as "ICR" (for *intelligent* character recognition), as opposed to the more limited OCR process. Some of the more powerful text scanning packages include the capability to output scanned files in the formats of various word processing packages, even to the extent of inserting the word processor's own commands for things like italics, underscoring, bolding, centering, and tabbing. Some also preserve multiple columns, or you may be offered the option of retaining or ignoring the columnar format of source documents.

To summarize this brief overview of text scanning, an OCR package is required to convert the scanned symbols into ASCII characters or into the format of your word processing software. If you want to exercise the latter option, before buying an OCR package be sure it supports your word processor. It's also important to keep in mind the kind of documents you will be scanning. If your needs are limited to typewritten or computer-generated source materials, you can save some money with an OCR package that uses the matrix-matching system for character recognition. But if you have to process typeset documents, be sure to get a product that performs feature extraction. Beginning on page 6, we'll be revisiting these processes in our discussions of scanning software products.

Picture Processing

When source materials consist of pictures or graphics, in scanner terminology we are dealing with *images*. You don't need a character recognition capability to scan images; to go back to our earlier analogy, image scanning software operates more like the camera. It makes a "copy" of the scanned page by creating a bit map of the page's contents. Remember, in bit mapping the file is made up of dots that are turned on (black) or off (white). Just as dot-matrix text is made up of different configurations of dot patterns, a bit-mapped graphic image is composed of millions of dots, each of which is or is not filled in. The more dense the dot pattern, the more numerous are the variations in shading that can be achieved. You could think of a scanned, bit-mapped image as a "snapshot" of the original hard-copy image.

It's important to understand these differences between text files and image files if you are concerned with the various purposes for which scanned files are used. For example, if you want to use a scanner to input raw text that will later be edited and imported into other documents (such as in desktop publishing applications), you should be aware that your source materials must be decent, but not necessarily perfect, and you need good OCR capabilities. On the other hand, if you simply want to use scanning to save documentation (that is, text that you won't ever need to edit again) in a more compact and convenient medium, you can process the pages of text as *images* without worrying about the quality of the source documents. The scanned images will capture the printed page like a picture, with all its tears, handwritten notes, coffee smears, and photocopy smudges intact—and it will be quite readable. Furthermore, there's no problem if the original document mixes text with photos, charts, and graphs; the image processing software sees all the elements on the page as parts of a single image.

Scanner-Generated Files

There are a lot of variations in front-end software for scanning text. The most basic products perform a simple matrix match on the scanned text and create an ASCII file, period. More sophisticated products, which will be discussed in more detail later in this report, come with software and/or firmware that speed up processing and have the capability to recognize a wide variety of fonts and prepare an output file in the format of any one of a number of popular word processing packages. File sizes for the ten test pages used in this study ranged from as little 3,500 bytes for a "normal" page of text to as much as 9K bytes for columns of numbers.

The Tagged Image File Format (TIFF or .TIF) file apparently is becoming the de facto standard for scanned image files. The most significant characteristic distinguishing TIFF files from text files is that image files can't be "edited" in the usual sense of the word. Often you can move a scanned image into paint program or a graphics package where you can move it around, alter its size, crop it, or rotate it. But if the file contains any text, you can't *edit* that text. Think of it again as a photograph. Once you've captured a photographic image on film you can alter it in some ways—darken or lighten it, remove parts of it, draw or write over portions of it—in the dark room. So you can modify the end product, but you can't really go back and *change* the original image.

A second, very significant, characteristic of TIFF files is their size: they are *LARGE*. A TIFF file containing one 8.5 x 11-inch page easily can (and often does) exceed a megabyte. Files containing complex graphs or pictures commonly are as large as 15 megabytes. The size of these files is a big stumbling block for lots of folks; many of us simply don't have enough memory and/or disk space to accommodate them. One solution, if the computer driving the scanner has enough memory to hold the scanned image and enough hard disk space to save it temporarily, is to immediately convert the TIFF file to another format before saving it. For example, we scanned a page, creating a TIFF file of around a megabyte; then used the WordPerfect graphics conversion utility to create a WordPerfect

graphics (.WPG) file that's only 218,000 bytes. It's highly probable that any loss of detail in the converted image will be noticeable only to the most critical observers.

Another thing to keep in mind that directly affects image file size is the *resolution* at which the image is scanned. For example, the same 1-page image scanned three times at 300 dpi, 150 dpi, and 75 dpi resulted in TIFF files of 65,754, 26,628, and 10,876 bytes, respectively. So if you can live with a lower resolution it can save a lot of disk space and speed up processing significantly.

Before we conclude our discussion of scanner files, it should be mentioned that disk files can be read and processed by most scanner front-end software and then be processed like input from the scanner itself. In other words, you can scan text or images today and save the scanned files on disk. Some time later, you can have the scanner software read the file from disk and process the image just as if it had come directly from the scanner. Text and images read from files created by facsimile (FAX) software can be processed like scanned images too. The capabilities of optical character recognition software can be particularly useful in this context. This will no doubt become more clear when you read the discussion of scanner software later in this report.

Product Evaluation Methodology

In keeping with PC TAP practice, users were heavily involved in this project. In addition to the TAP staff and our colleagues in the information centers at Research Triangle Park, participants from several other RTP offices, the Washington Information Center, Regions IV and VIII, and NEIC were active in the study. Thirteen scanners and eight software products were evaluated.

When we devised our evaluation materials, we didn't make it easy for the scanners. Folks who knew about our scanner study and who are interested in exploring scanning technology brought materials for us to use. "See if you can scan *this*" was commonly heard. Often these source materials represented a real challenge, because they definitely weren't "crisp" copies. Apparently there are a number of folks who have only hard copy (frequently mountains of it) of data they want to use, but for which the original computer files have been lost. These people see scanning as the solution to their dilemma. Just scan the hard copies to restore the data files! Certainly it's a possibility, but the condition of the available source documents is the key to the viability of the scanning solution. Some of the scanners and software we've looked at are very good, but they aren't *magic*; even great technology can't do a satisfactory job with 5-year-old 3rd or 4th generation photocopies of reduced laser printer hardcopy output. But we tried.

Our evaluation packet included ten pages of source documents that we asked participants to scan on their equipment: a typical image (the cover page from a training manual); mixed text and images (pages from technical manuals containing text along with scientific notation, tables, and pictures); and text pages containing typewriter-like type faces, typeset material (including multi-column pages and mixed fonts on a page), computer-generated tables, and straight text in both a typewriter-like face and a non-typewriter font from a PC word processing package. Study participants were asked to save the scanned files on a floppy disk provided with the evaluation materials and return it to PC TAP. They also completed a questionnaire on which information about their scanning hardware/software was recorded along with their evaluation of its performance.

We have elected to discuss the various software products that were included in our study first. An overview of each product is presented in the next section. Then in the hardware product reviews beginning on page 12, we will discuss each scanner's performance in terms of the front-end software that was used for the tests.

Product Reviews: Software

One should consider several key points when selecting an OCR product. The first is hardware compatibility. It doesn't matter what the software will do, if you can't run it on your system it's worthless as far as you're concerned. Hardware compatibility turns out to be a bigger potential barrier than we would have guessed. First you have to be sure the software will run on your computer (e.g., MS-DOS vs. Mac). We discovered a lot more scanning products for the MS-DOS environment than for the Macintosh user, but the gap seems to be closing. You also have to be very careful to ensure that your scanner is supported by the software. All OCR products are not compatible with all scanners. In summary, there are three links in the scanning chain: (1) the scanner itself, (2) the computer to which it's connected, and (3) the software for processing scanned text and images. When you're putting together a system to do scanning, all three links must be mutually compatible.

Performance factors related to OCR software include speed, number and types of fonts supported, text recognition accuracy, and supported file types. The text-recognition process is an involved one, and it can take considerable time. Essentially the software has to look at each character in the file and make a decision about what that character is. This process is usually accomplished by comparing the characters in the scanned file to character tables that are part of the software. Some products are more efficient at this process than others, resulting in measurable differences in the time it takes to "recognize" a page of text. Reported scan/recognition times for devices in our study ranged from 30 seconds for straight text to as much as six minutes for complex pages (mixed text/graphics, multi-columns, "hard-to-read" copy).

We made reference earlier to two different methods of text recognition, matrix matching and feature extraction, and pointed out the characteristics of each. OCR software may operate by either of these methods; some products use both. The flexibility of the product is reflected in its text-recognition capabilities, and it's important to remember that the font recognition capabilities of a package that uses only matrix matching will be limited. You have to be careful, too, in interpreting accuracy claims of software vendors. In their advertisements they often say their product averages "98 percent accuracy" (or some other number approaching 100%) in tests of text recognition. This may mean that the software was unable to even make a guess at two percent of the characters it encountered. It *doesn't* necessarily mean that the software *correctly* identified the other 98%—just that it "thought" it did.

Finally, the number and types of files supported by an OCR package are an important measure of its performance. Some only output ASCII files. If you want to use those files with a word processor or desktop publishing package you have to import them and edit them accordingly. The more sophisticated products will produce files in the format of any of a number of word processing packages. You simply indicate the package you want to use, and a file in the proper format—including formatting codes—is generated.

In the following paragraphs software and firmware products are presented in alphabetic order by product name. No quality ranking should be inferred by the order in which these products are discussed. To refresh your memory, the term *firmware* is applied to processing instructions or programs that are contained on a microchip, rather than in memory or in a disk file. PC scanning products often come with boards on which the OCR software resides on a microchip, along with memory chips that help speed up processing.

AccuText

AccuText is an intelligent character recognition package from Xerox Imaging Systems. It processes both images and text. According to the AccuText literature, it is capable of recognizing "thousands of

type styles in sizes ranging from 8- to 24-point on both portrait and landscape pages." The product is advertised to recognize typeset, laser printed, impact printed, typewritten, and letter-quality dot matrix printed pages. It also has a built-in 50,000-word dictionary and context rules, so it checks the spelling and structure of the source materials during the character-recognition process. In addition, a user dictionary can be created with up to 10,000 special terms that also will be checked. Text in multi-column format can be read successfully. Output files can be in Microsoft Word RTF, Microsoft Excel, Claris MacWrite, or text-only format.

AccuText supports image scanning in resolutions of from 60 to 450 dots per inch, depending on the scanner in use. Scanned images can be output in these formats: TIFF Uncompressed, TIFF PackBits, TIFF CCITT-3, PICT, and MacPaint. A "Preview" command allows you to preview a scanned page and identify text and image areas and specify the order in which they are to be processed. Areas that are not to be scanned may also be identified. You also can choose whether to process text and images separately or in one step.

We weren't able to test a production version of AccuText, but we did obtain a demonstration version for one of our study participants who's on the market for a Macintosh OCR package. Our evaluator didn't think the software lived up to its press, but the demo package was severely restrictive and did not permit all AccuText's features to be tested. With regard to text recognition, results from scanning our ten test pages were encouraging. Several did very well, but others were totally unsatisfactory. Macintosh users who are looking for a character recognition package would probably be well advised to explore a production version of AccuText more carefully.

Discover 7320

This software was bundled with an older Kurzweil Discover 7320 Scanner. It's a text-recognition package that uses ICR technology to recognize typewritten, laser printer, and typeset materials. Dot matrix hard copy is not supported. Compared to the other software products in our study this one is older, and it has one capability that the newer ICR products no longer need: it's *trainable*. This means you can literally sit down at the computer and, by describing the characteristics of the characters, "teach" the software to recognize a font. Although we've never tried this task, everything we've read or heard indicates that it's a long, painstaking, tedious process. More recent products like Accutext, OmniPage, and TrueScan have the built-in capability to "learn" fonts without human intervention. The Discover software will process scanned pages in either landscape or portrait orientation, and the original document format is preserved. ASCII is the only supported output file format.

Although our evaluator reported reliable text recognition performance at acceptable speeds, newer and more sophisticated products are currently available. Users interested in Kurzweil scanners and software should be aware that Kurzweil has become part of Xerox Imaging Systems.

OCR Plus

OCR Plus is a "third party" product that's shipped with several manufacturers' scanners. Input we received relative to use of OCR Plus was in conjunction with Datacopy Model 200 and 320A scanners in the MS-DOS environment.

For character recognition, this product uses matrix matching "supplemented by a topological technique." Like the Discover software described above, it's trainable when you need to scan fonts that aren't built in to its character-recognizing repertoire. When using OCR Plus in conjunction with tests of the Datacopy 730GS scanner, *PC* magazine reported performance "on a par with other scanners" in tests limited to 10-point Courier type. However, less success was achieved with proportional fonts and mixed type sizes.

Our evaluator's comments support PC's findings. While recognition accuracy was acceptable with the 10 or 15 fonts OCR Plus "knows," the best that was achieved with typeset material was "probably 75 percent accuracy." Overall, the best text-scanning results were achieved with documents printed on laser printers and from a 24-pin dot matrix printer with a new ribbon. Our study participant "taught" OCR Plus a font, and reported that the process took a great deal of time. During the "teaching" process, letters had to be typed in with no errors. There was no way to edit a character after it was entered, so if a mistake was made it was necessary to recreate the file and start over.

OmniPage

Caere Corporation's OmniPage is a first-class product. We tested version 2.0 on both a Macintosh II and an Epson Equity III+. The MS-DOS version, which comes with software and a companion board that takes up a full slot in the PC, is designed to run under MS Windows. In case you don't have Windows on your computer a run-time version is bundled with OmniPage. The Mac version needs no board or Windows interface. Just load the software; it looks and acts like the typical mouse-driven Macintosh application.

When you install OmniPage you are given the opportunity to set a number of default options for output files, including selection of the format for text files from a list of supported word processing packages. However, each time you scan a document you have the option of overriding one or more defaults, so there's plenty of flexibility built in to the product.

OmniPage gives the user a lot of visual feedback, along with meaningful messages about what's going on during the sometimes lengthy (30-120 seconds, depending on page complexity and scanner options selected) scanning/text-recognition process. In addition, while text-recognition is going on, a small window is opened on the screen in which characters are shown "as the software sees them," giving the user some feedback about how well the source document scanned, and whether using the "lighten" or "darken" options might improve recognition. Visitors to our information center really liked these features. There is a quick scan option that reads a page into a temporary file that you can then look at to see whether you want to make any adjustments to contrast or other mode settings before proceeding. Once you're satisfied, you can select the normal scanning mode to process the current page and any more that follow. Settings established for the first page in a multi-page operation are retained throughout the session unless you change them.

OmniPage is an *omnifont* product: it can read a wide variety of fonts, and handles type sizes of from 8 to 72 points. Multiple columns are accommodated, as are source documents in both portrait and landscape orientations. A partial page option allows you to define a specific area of the page to be recognized, while the rest of the page is ignored. We found we could narrow this area down to a single word with no trouble. Character recognition speed is advertised as from 40 to 115 characters per second. Unrecognized characters can, at the user's option, be "flagged." The tilde symbol (~) is placed above questionable characters in the text file when the "show suspects" option is turned on.

Although OmniPage supports a number of scanners, some are not included in its list of supported devices. However, there's a way around this problem too. Simply scan a page of text into a TIFF file ("take a picture" of the page), then read the resultant file with OmniPage's "Recognize" command. The text in the TIFF file is "read" by the intelligent character recognition software, and a text file in the format of the selected word processing package is created.

Release 2.1 of OmniPage, for Macintosh II's and 386 and 486 PC's, was announced by Caere Corporation in November. It will read and write both compressed and uncompressed TIFF files (version 2.0 only handles uncompressed TIFF files), and has the capability to interface with a number of companion products like Omnispell (a spell checker) and Omnidraft (recognizes dot-matrix fonts).

Although we haven't had an opportunity to try release 2.1, we were very pleased with OmniPage 2.0 and can recommend it highly. More discussion of OmniPage can be found in the section describing our tests of the Hewlett Packard ScanJet Plus scanner.

Publish Pac

Publish Pac is a desktop publishing package designed for use with IBM XT, AT, and PS/2 computers (and compatibles) and any of the DEST PC Scan series scanners. It runs under Microsoft Windows, and a run-time version is included with the Publish Pac software. A graphics adapter card and a mouse are required. The documentation that's provided with the software was judged "better than average" by our evaluator.

This product has a good user interface, with pull-down menus and easy-to-understand messages. Our evaluator particularly liked Publish Pac for scanning images, as opposed to text. When you don't need the entire contents of a source document, it's easy to identify a particular part of the image to be processed. After the scanned image is displayed on the screen, you just use the mouse to "draw a box" around the selected area, and click OK when you're satisfied. The portion of the image inside the box is all that will be placed into the file created by Publish Pac. Image files can be saved in any of four formats: TIFF (.TIF), PC Paintbrush (PCX), uncompressed (.IMG), and Encapsulated PostScript (.EPS).

The text processing capabilities of Publish Pac are somewhat limited. Only typewriter-like characters and a few fonts from laser printers are recognized, and unrecognizable characters will be represented in the scanned file by the pound symbol (#). In addition to standard alphanumeric characters, only a limited number of special characters (* \$ # @ / () & - + = £) will be recognized. This means Publish Pac will not be a satisfactory product for people who anticipate a requirement for scanning typeset source materials. Text files may be saved only in ASCII format.

On the plus side, Publish Pac has the capability to scan images and text together. After the scan operation is complete, you can create an ASCII file into which the text portion is saved, and an image file containing the graphic portion of the page. The image file can be in any of the supported file types listed above. Publish Pac was used in conjunction with our evaluation of the DEST PC Scan 2000 and DEST PC Scan Plus scanners.

ReadRight

ReadRight is an OCR product that's bundled with the Hewlett Packard ScanJet Plus and several other manufacturers' scanners. Our copy says it's designed to be used exclusively with the ScanJet; an HP ScanJet interface card is required. It is compatible only with version 3.0 or higher of MS-DOS.

The documentation, which is excellent, says it's the "first low-cost high-performance topological OCR system." *Topological* is another way of saying "feature extraction." This sounds great until you find out that the only fonts that ReadRight recognizes with this technique are the typewriter-like character sets. The result is very good character recognition accuracy, but with a limited number of fonts. Specifically, nine "monospaced" (all characters, including spaces, take up the same amount of horizontal space in the line) and ten "proportionally spaced" (characters take up unequal linear space) fonts are listed. In the ReadRight manual, under "limitations," it says the product can't yet read "typeset documents, documents printed by a loose dot-matrix printer, and poor photocopies."

ReadRight has the usual options for controlling contrast (they call it print intensity), scanning resolution, and paper size of the source document (8.5 width, 11-14 inches length). There's also an option to have the text file written directly to a disk file without displaying it on the screen. This option speeds up processing, but obviously you can't monitor what's going on or check on the accuracy of text-

recognition. Output files can be in any of three formats: ASCII, WordStar, or WordPerfect. In addition, there are three versions of ASCII. The first, called *ASCII WP*, puts only one space after each word (even if the original had two), inserts a carriage return at the end of each line, and inserts two spaces after a period. The second, *ASCII DTP*, puts a space after every word (even if the original had two or more), puts carriage returns only at the end of paragraphs—not at the end of each line. Finally, *ASCII WYSIWYG* reproduces the document in its original form using only spaces and carriage returns, but no tabs.

In our tests of ReadRight with our HP ScanJet Plus, we found it to be very accurate in scanning the fonts it "knows." However, nothing usable resulted from scanning anything but typewriter fonts during our evaluation.

Scanning Gallery Plus

Hewlett Packard bundled this image-scanning product with the HP ScanJet Plus scanner. It runs under Microsoft Windows, and a mouse is required. When Scanning Gallery Plus is started, two windows are presented on the screen. The *Scanner* window is where the user engages in a dialog about the scanning operation. Here you can specify the type of scanning operation you want to perform, adjust the contrast, ask for a "preview" scan, indicate that just a partial area of the source document is to be processed, set the dimensions of the image to be saved in the TIFF file that will be created, and name and save those files. The second window, the *Image Editor*, is where you view the scanned image and select partial areas to be processed if you wish.

Scanning Gallery Plus comes with excellent user documentation that gives detailed instructions about the use of the various options offered on the scanning menu. Gray scales are supported, and the user can select from among four dithering patterns for photographs. A utility is provided to convert Scanning Gallery Plus' standard TIFF files to MSPaint, PC PaintBrush, GEM, or Encapsulated PostScript files. An editing feature allows cutting, pasting, and cropping of all or part of an image.

We found this product easy to learn and use. Compared to some other products that offer scanning of partial images, it's easy in Scanning Gallery Plus to indicate the portion of the image you want to process: you just use the mouse to draw a box around it. Repositioning and cropping of image elements is equally quick and easy with the cut-and-paste function. For image scanning, this software is all most users of Hewlett Packard scanners should need.

TrueScan

TrueScan was honored by *Byte* magazine with a 1989 "BYTE Award of Excellence." These awards are given to products deemed to be the year's most significant new offerings, and that are the personal favorites of *Byte* editors and columnists. Additionally, *PC* magazine called TrueScan "a powerhouse" product. A shortcoming in the minds of Macintosh users, however, is that it's only available for MS-DOS machines.

Like OmniPage, which we discussed earlier, Calera Recognition Systems' Truescan comes with both software and a board. One unique feature of Truescan, however, is that an optional "daughtercard" that can piggy-back onto the controller boards of some (but not all) scanners, thus saving a slot on the PC. Performance is said to be "about ten percent better" if you choose the daughtercard rather than a full Calera board, which is also available.

Calera offers a whole range of scanning products. TrueScan is available in two models for PC/AT's and PS/2's and compatibles, Model S at \$2795 list and Model E at \$3995 list. Model S scans at speeds of up to 75 characters per second and reads only in portrait orientation. Model E operates at speeds

of up to 100 cps, and handles portrait, landscape, and rotated pages (FAX images). We tested the Model E, and found its performance lives up to its publicity in most cases.

TrueScan's list of supported scanners and word processing packages is impressive, and much too long to list here. Suffice it to say that chances are excellent that your word processor will be supported; that is, files in the word processor's format can be generated from scanned pages. The list of supported scanners isn't quite so comprehensive, but most of the front-runners are included. A wide variety of output formats for images is supported too, and scanned tabular information can be plugged into Excel, Lotus, and Quatro spreadsheets.

We tested the full-board (no daughtercard) version of TrueScan Model E with our HP LaserJet Plus. Results were excellent. Our only negative criticism relates to the user interface. We didn't find this product as user friendly as OmniPage. There is very little visual feedback, and some of the status messages are cryptic and not totally accurate. For example, the scanning and text-recognition processes are two separate steps in the overall process. TrueScan presents a "Scanning" message when the light comes on in the scanner and the process begins. That initial message remains on the screen with no changes or status updates while the scanner light goes off and the PC goes to work on the text-recognition process. If you understand what's going on, it's not so bad; but when we first started using the product we were baffled by the "Scanning" status message that remained on the screen long after the scanner obviously had finished doing its job.

Overall, it's hard to fault TrueScan's performance. According to Calera, it can recognize over 16,000 fonts (some of which must be variants of the same basic type face); character recognition accuracy with good source materials is said to be as high as 99.9%; both text and graphics are captured in one pass through the scanner—text goes into the user-specified word processor file, graphics into an image file; multiple fonts and/or type sizes on the same page are handled with ease; and a built-in spell checker flags misspelled words as well as doubtful or unreadable characters. In the low-end class, TrueScan is the most powerful product of its kind that we've seen—but it's the most expensive too.

Summary

As is usually the case when you look at a lot of different software that is designed for the same application, there are a lot of similarities among the products in our study. Just about all image scanning and OCR packages currently on the market live up to their manufacturers' claims pretty well. Certainly the ones we looked at did. They key, then, is to look at what's *claimed* for a given package, and make sure it's suited to your purposes.

First and foremost, the software must be compatible with your scanner/computer configuration. Be sure also to check the OCR/ICR capabilities if you're planning to do a lot of text scanning, and verify that the product will produce an output file your word processing package will handle with ease. The format of scanned files is also important with respect your image scanning needs, so check for compatibility of those files with software you intend to use for modifying and printing scanned images.

The ultimate criterion for many of us when it comes to selecting software for any application is cost. Just as the products in our study have diverse capabilities, they also represent a wide price range. Some basic, software-only OCR products start in the \$500-\$600 range; the True Scan Model E we tested lists for \$3995. So look at your potential scanning needs to get a handle on what functions the software must support, find products that will run with your hardware configuration, and choose the best you can afford from among the packages you've identified.

Product Reviews: Hardware

Each of the scanners evaluated in our study is discussed in the following paragraphs. No ranking is intended by the order in which they are discussed; the devices are presented in alphabetetic order by product name. A table summarizing the features of all the devices we tested appears on page 20.

Scanner Devices

Before discussing the particulars of each individual scanner, it will be helpful to briefly review the capabilities and features of scanners in general. Fundamentally, they all work on the same principle: light is bounced off the source document, and the scanner measures how much is reflected back. The reflected light generates a variable amount of voltage in a sensor; the more light that comes back, the higher the voltage. Zero voltage translates to black, and increasing voltage generates ever lighter shades until the highest voltage yields white. One aspect in which scanners are judged is the number of shades of gray they are capable of producing. Some are capable of only 2 levels (black and white), while the better low-end devices can distinguish 256 shades of gray. Since the reflected light patterns are used to create the bit maps we discussed earlier (see *Picture Processing*, p. 4), the greater the device's capability for gray-scale recognition, the finer the bit-maps (and the larger the files) it will produce.

When it's time to produce a hard copy of a scanned image, it doesn't matter how good the scanning software is if resolution of the output device isn't compatible with that of the image. Resolution is a product of the density of the bit-mapped dot patterns discussed earlier; denser patterns accommodate more shades of gray, yielding higher resolution. Excellent results can be achieved with a scanner capable of 300-dot-per-inch (DPI) resolution and 256 shades of gray, and a 300-dpi PostScript laser printer. It's worth mentioning again, however, that very large files are required to accommodate images with these characteristics. Two methods are employed in software to achieve gray-scaling in scanned images. The first is *dithering*, a process by which the density of the bit map is altered before the scanned file is saved. The dithering, then, is stored with the image. The second, more recently-developed technique is called *gray scaling*. In gray scaling, values representing the gray tones (rather than bit patterns) are stored with the image. Creation of the pattern occurs when the image is sent to the output device, so the software tailors the output to the capabilities of the printer. The TIFF files mentioned earlier are the most common format in which gray scale images are saved.

There are two basic physical configurations for scanners, flatbed and sheetfed. Flatbed scanners resemble photocopy machines (except that they're usually a lot smaller). You lift a cover from the glass surface, place the source document face down on the glass, close the cover and start the scanning operation. The light source inside the device passes beneath the source document and does its light-bouncing job, the image is captured, and that's that. With sheet-fed scanners, the source document usually is fed between rollers that "grab" the paper and feed it through the inside of the device where the scanning operation takes place. The source document is then returned to the operator through an opening at the end of the device's "paper path." In both cases, you give the machine one page at a time, unless you purchase an optional document feeder (available with some scanners) that accepts a stack of documents that are automatically fed to the device one at a time. One disadvantage of the sheetfed scanner is that you can't lay an open book on the glass to copy a page; nor will it accept thick materials. As the name implies, sheetfed scanners accommodate one sheet of paper at a time. Period. Sheetfed scanners also have a reputation for jamming source pages in the paper path. Flatbed scanners, on the other hand, will handle both the open book and other heavier-than-paper source materials.

Handhelds

We've said there are two basic scanner types, but a third type deserves mention here: hand-held scanners. We didn't include any hand-held devices in our study. Our task was defined as "evaluating desktop scanners." Nevertheless, during our research we came across some information about hand-held scanners, and we considered trying to find some we could test. However, the negative feedback we got from people who already had looked at them led us to dismiss the idea. Many people feel that good handheld scanners will be available some time, but they aren't here yet.

For our readers who are interested in hand-held devices, here's what we know in a nutshell. The Mitsubishi Handheld Image Scanner (no text recognition capabilities at present) is currently available at a list price of \$995. An optional sheet-feed attachment, to which the scanning device quickly attaches to make a flatbed desktop unit, costs another \$260. In hand-held operation, this device is said to do an acceptable image-scanning job, but lack of a text-scanning capability puts it out of contention for most scanning applications we've been confronted with by EPA users.

Another hand-held image-only scanner we read about is "ScanMan" from Lotus Selects (PC version \$339 list; PS/2 version \$399). ScanMan has a 4-inch scanning window that allows you to scan images up to 4 inches wide and 11 inches long. Images can be scanned into TIFF or PC Paintbrush format, and can be saved into TIFF, PC Paintbrush, or Microsoft Paint format.

When we were researching the literature in preparation for our scanner project, we found a somewhat dated review (*PC Magazine*, Jan. 26, 1988), of the *Complete Hand Scanner* from Complete PC Inc. The device offers 200-dpi resolution and a 2.5x10-inch scan path for \$249. It was said to be "very good" for black-and-white line drawings, while photographs were "more challenging." The front-end software converts images to Dr Halo, PC Paintbrush, and Windows formats. A "bad manual" was pointed out as the primary shortcoming of the product. Like most other hand-helds, no text scanning is supported.

Along with the input provided by one of our study participants was an account of one site's local assessment of handheld scanners from Logitec. The device is limited to a 4.5 x 6-inch scan, and getting it properly aligned for text scanning was said to be a problem. (Text alignment in even the better flatbed devices is critical; the text on the printed page needs to be perpendicular to the path of the scanning wand--except, of course, in the case of landscape orientation.) Scan speeds were said to be slow. Our evaluator summed up this device as "an OK toy."

Now that you've had a quick primer on scanners, let's look at the individual devices. Evaluation data for these narratives was provided by the participants in our scanner assessment project. For some devices general evaluation material and user comments were received, but data on scanning the test documents were not included. In those cases, only the available general information is summarized. When detailed test data is included in the discussion of a particular scanner, that information was provided by the participants who actually ran the tests on their respective equipment.

Apple Scanner

As sometimes happens with PC TAP studies, the person from whom we expected an assessment of the Apple Scanner was unable to complete the study. However, we feel this product deserves mention in our report, so we're including a summary here of some general information that appeared in several trade journals.

The Apple Scanner is a flatbed model offering resolution of up to 300 dots per inch when processing line art, photographs, and gray-scale images. One shortcoming is a limitation to only 16 shades of

gray, however. The scanner is a SCSI device, so it works with any Mac Plus, SE, or Mac II that has System Version 6.0 or later.

Both AppleScan and HyperScan software come with the Apple Scanner. These packages provide for scanning (directly into HyperCard stacks if you choose), cropping, sizing, and fine-tuning images. Source documents in both landscape and portrait orientations are accepted. For text scanning, OmniPage supports the Apple Scanner, and is reportedly a popular ICR product among Macintosh users. We have seen the retail price for the Apple Scanner reported at both \$1699 and \$1799.

Chinon Desktop Scanner

The Chinon Desktop used in our evaluation was an older model. It's a serial device, and is slow in operation. Scanned image files were moved into Chinon graphics software for further processing. These images had good resolution (although images with lots of arcs and diagonal lines were avoided), and it was possible to size the image within the graphics package.

A recent Chinon scanner, the DS-3000, was favorably reviewed in the March 28, 1989 issue of *PC* magazine. This device, classified as a "portable" scanner, is intended for the desktop publishing market. At \$745 it comes with bundled image-processing software. For \$995 you can buy the DS-3000 with an image-scanning utility and ReadRight bundled in (see page 9 for more about ReadRight).

The DS-3000 has a unique characteristic: It's an *overhead* scanner. It looks a lot like a portable overhead projector. You lay the source document on a flat bed, and the light source is housed directly over it atop an arm extending from the back of the scanner. In the *PC* review of this product, they said that because the source document is virtually unprotected from external lighting effects, all their tests yielded images in which shadowing effects were present. They placed heavy emphasis on portability and desktop publishing applications, but this scanner's suitability for general office use was left open to question.

Datacopy Models 200 and 320A

We didn't receive any detailed evaluation data about the Datacopy Models 200 and 320A. These devices were used in some local scanner tests at one of our participating locations, and the results of those tests were forwarded to us. However, our ten standard test documents weren't included in the local tests, and no assessment of how our tests fared on these devices was included in the information we received.

Document scanning done on these devices was accomplished with the aid of OCR Plus, which was discussed on page 7. Scan speed was characterized as "slow." Reasonable text recognition accuracy was reported when source documents were of good quality ("not a copy of a copy of a ...") and the font was one the OCR software could "read." In some cases, the success rate of character recognition was improved by enlarging or reducing source documents on a photocopier in an attempt to approximate a recognizable font. It was reported that "almost anything that was (typeset) ... could not be satisfactorily scanned."

Datacopy Model 830

Our evaluator with the Datacopy Model 830 scanner is a Macintosh user. Although this is an excellent scanner (it was rated "best for Macintosh users" in a 1988 review by *Publish!* magazine), our study participant has had difficulty finding suitable front-end software to use with the device. Although a lot of hardware still bears the Datacopy name, the company is now a subsidiary of Xerox Imaging Systems.

For purposes of completing our scanner evaluation, this participant used a demonstration copy of AccuText, a Xerox Imaging Systems product for the Mac. Given the limitations imposed by the demo package, this software performed quite credibly. Some formatting problems were encountered, but this is common in scanned documents. A lot depends on how the scanner was set up, for example specifying multiple columns or landscape oriented material, before the operation was begun. Despite the sometimes strange appearance of the scanned files, a careful reading of the text reveals a very high level of character recognition accuracy.

The Datacopy Model 830/AccuText rendering of one particular page that was the "acid test" that most of the OCR software in our study failed is very good (a rather poor photocopy of many columns of numbers in a small typeface). It would probably be acceptable for production work as a viable alternative to re-creating the source material from scratch. As we said in our software review of AccuText, this combination looks like a viable option. However, we recommend a more careful evaluation with the production software before making a decision to purchase.

DEST PC Scan 2000

This device is compatible with both IBM PCs (and compatibles) and Apple Macintosh computers. Our evaluation device was attached to an IBM PC/AT, requiring installation of a scanner interface board in the computer. Scanning of both images and text is supported, the latter with the bundled Publish Pac software. An automatic document feeder (ADF) is available as an option, but the device used in our evaluation didn't have this attachment. However, with the installation of a FAX board in the computer the scanning station has been used quite successfully as a FAX terminal as well.

The PC Scan 2000 is a sheetfed scanner, and the biggest physical complaint about the device is its inclination toward crooked paper feeding and jams in the paper path. Frequent users claim the odds of an improper feed are greater than those for success. Additionally, the availability of more sophisticated text-recognition software has been accompanied by a sharp decrease in demand for this device as a text scanner. Our tests were conducted with Publish Pac as the recognition software (see discussion under "Product Evaluations: Software"). Nevertheless, our evaluator did give the PC Scan 2000 high marks as an image scanner (with a caveat for the troublesome paper-feed characteristics).

DEST PC Scan Plus

The DEST PC Scan Plus came bundled with Publish Pac software by Silicon Beach. This product doesn't read dot matrix source materials, but it does handle output from typewriters and laser printers, along with typeset documents. Only source documents in portrait orientation are accommodated.

Our evaluator, who uses the PC Scan Plus with a Macintosh, reported better results with scanned images than with text. Accuracy of text recognition seemed to be fairly font-specific; clear copies of some type families were scanned with low recognition accuracy. The documentation for both the hardware and bundled software were rated "average." Speed of operation was said to be unacceptable.

In processing our test pages, the PC Scan Plus performed about as expected with the configuration described above. The typewriter fonts were read fairly accurately, with the Prestige Elite coming out better than the Courier. The typeset pages were worthless. Image processing was quite good, and zeroing in on one field on a travel voucher was excellent.

Commenting on the most-liked features of the DEST PC Scan Plus, our evaluator listed "easy-to-use front-end." Things liked least included "sheet feed limits paper size; no magazines, books, etc.; pulls

paper crooked frequently.' It was noted that this device is several years old, and better products have become available more recently. With this in mind, readers who are looking for a scanner to purchase are advised to look at other products.

DEST Workless Station Model 202

The DEST Workless Station is a standalone text scanner with built-in firmware that produces an ASCII file. Typewritten character sets and output from laser printers in typewriter fonts can be read, but no dot matrix or typeset material is recognized. The device has no graphics scanning capability, and reads only in portrait orientation. This is an "older" scanner; it cost around \$10,000 in 1985.

The biggest objection to this scanner is that, rather than connecting directly to the computer, it requires an ASCII communications connection to the serial port in the PC. Robert Root, an IC consultant at the Washington Information Center, reported to us on the DEST Model 202. His concise description of the device is so comprehensive that we reproduce it here:

The DEST Workless Station Model 202 is the most reliable, mechanically and electronically, of the four scanners we have. It is also the simplest to use because of its reliable document feeder and its two controls: a button to "read" and a button to "clear" if the operator wishes to cancel scanning on the current page. The only complexity results from having to know how to tell the PC software, Crosstalk XVI in our setup, how to capture and save a specific ASCII file on disk. Scanned ASCII text is transferred to the PC via a serial port at 1200 bits per second during and after the page scan, so large stacks of pages processes quickly and efficiently.

The red illumination at the scanning window permits use of black type to fill in preprinted orange or red ink forms so that only the filled-in contents of the form are read. This feature could be a real time and error saver for certain data entry applications, but to my knowledge has not been exploited during the 5 years we have offered this scanner to EPA headquarters users. It is a real shame that our more modern and capable scanners don't have as simple a user interface. I see little reason why they couldn't.

Our ten test documents were scanned on the Workless Station with mixed results. Understandably, images and symbols were not properly recognized. Text recognition accuracy for pages containing text in typewriter fonts ranged from good to excellent, and photocopying the "originals" (which were in fact photocopies in the first place) to darken the text and thicken the characters resulted in improved scanning accuracy in some cases. (It was noted on the evaluation form that "copies must be high quality for good scanning accuracy.") We must point out, however, that tests with today's ICR software yielded equal or greater accuracy with no "enhancement" of source documents.

Hewlett Packard ScanJet Plus

The PC TAP staff have access to a new HP Scanjet Plus in the information center at RTP. We did extensive testing with this device on both an Epson Equity III+ and a Macintosh II. In the MS-DOS environment we used HP Scanning Gallery Plus, ReadRight, TrueScan, and OmniPage software to process scanned files; overviews of these products are in the section of this report dealing with software.

The ScanJet Plus is a flatbed scanner. It comes with a board that must be installed in the PC before you can use the scanner; a board is not required for the Macintosh. For MS-DOS machines, the scanner is shipped with two software products: the HP Scanning Gallery for image scanning, and ReadRight, an OCR product. Scanning Gallery Plus, which runs under Microsoft Windows, handles source images in both portrait and landscape formats. If your machine doesn't have Windows, a run-time version comes with the HP software. Both Scanning Gallery Plus and ReadRight are mouse-driven and easy to use. If you're anti-mouse, you can still use the keyboard to run the software.

Details of our experiences using TrueScan and OmniPage with the ScanJet Plus may be found in the discussions of scanning software. Retail list price for the ScanJet Plus is around \$2,000.

We have been very pleased with the performance of our scanner. It's easy to operate, has no confusing or cumbersome knobs or switches, and has been trouble-free in both the PC and Macintosh environments. Clients in our information center have little trouble using it, and they invariably are pleased with the results when they know how to use the scanning software properly. We can give an unqualified endorsement to this device.

On the Macintosh II we used OmniPage to scan our ten test documents on the ScanJet Plus. OmniPage, ReadRight, and TrueScan were all tried on an Epson Equity III+. An advantage of the Mac over the standard AT-class PC for scanning is that there's no need for adding a board to the computer. Once the image has been captured, though, it's more a matter of user preference for the working environment. We didn't notice any appreciable difference in the quality of text or images that we could identify as CPU-specific.

Kurzweil Model 4000

Like the DEST Workless Station, the Kurzweil Model 4000 is a "stand alone" scanner that must be accessed through a communications interface. Reflecting another similarity to the DEST, our study participant used Crosstalk to address the scanner. The Model 4000 is a "text-only" scanner with no capability to process images. All scanned text is saved in ASCII files. This configuration was characterized as "old," and since more direct connectivity is available with newer products, the Model 4000 is not recommended for individuals currently looking for a scanner.

The success of this device in reading our test files is a testimony to Kurzweil's reputation as a leader in the scanning industry. Even its "old" technology demonstrated excellent character recognition capabilities. Although it did have trouble with a couple of pages, for the most part a very high reliability was demonstrated. This product did an outstanding job with the "hard to read" columns of numbers.

Kurzweil Model 7320

The Kurzweil Model 7320 with OCR software and coprocessor board was a \$10,000 investment when it was purchased in 1987. A subsequent upgrade for the OCR software in April 1989 cost an additional \$400.

The study participant who reported on this product cited no problems installing or using any part of this configuration. However, the document feeder has been a chronic irritant after the first 25-50 hours of service. It requires constant monitoring because of a tendency to "grab" several pages at a time. Another disliked feature is the "complex, menu-driven user interface that can't be bypassed or streamlined for simple production scanning of multi-page text documents unless the pages feed reliably."

In a more positive light, the 7320 was reported to have a very flexible font-recognition capability. In addition, the capability of fine-tuning scanner and OCR settings from on-screen menus was seen as a significant advantage. Although the performance of this scanner was rated highly, because of its troublesome document feeder and cumbersome user interface, our evaluator did not recommend that others consider acquiring a similar device for their office use.

This scanner turned in a top-notch character-recognition performance in processing our test documents. It rates among the top of the group. Regardless of font, text pages were reproduced with few or no errors. Sometimes formatting was not totally maintained, but it wouldn't require a major effort to remedy

the discrepancies. Like the Kurzweil 4000 discussed above, this scanner did an excellent job on the columns of numbers that were troublesome to many of the other devices.

Microtek MSF 300A

The Microtek 300A is a flatbed scanner which, according to reports in the literature, is a first-class device. However, the report from our evaluator didn't include a recommendation that other users consider acquiring one. Although some hardware incompatibilities were encountered when the scanner was acquired, no significant operational problems were reported with the device. But our evaluators' experiences have not generated much enthusiasm for using it. Scanning performance was said to be "fine," but slow, and the scanner itself was rated "okay."

This was a field-tested scanner, and we have no first-hand experience with either the device or the front-end software that was used during the testing. The image-scanning software is a product called Eyestar Plus; SmartStart was used for text. Neither was rated satisfactory by our study participant. The text-recognition software was said to work "fine with simple text, but is not very flexible." This sounds like what you would expect from a matrix-matching product; with fonts it "knows" it does an acceptable job, but otherwise performance is limited. The image-scanning product was summarized in this way: "works for scanning pictures as long as they are very sharp."

When our test pages were scanned on the 300A, the results were for the most part unusable. Although some pages (not surprisingly the typewriter-like source materials) scanned better than others, even the best weren't suitable for production work. A good typist could re-enter the text in less time than it would take to edit the recognition errors out of the scanned files. In some cases, practically nothing of the source text was recognizable.

The image file that was to have contained the picture of the factory only held the title line from the page on which the picture appeared on the original document. We suspect a memory or file-storage limitation caused this. However, when the software failed to produce a file from two of the text pages, our study participant scanned those pages as images. This resulted in quite readable (but un-editable) images of the original text.

Overall, our test results support the evaluator's less-than-enthusiastic endorsement of the Microtek 300A. Based on our experience to date, however, we suspect the lackluster performance may be attributable more to the image- and text-processing software than to the scanner itself.

Microtek MSF 300G

This device was evaluated in the Macintosh environment using Microtek DA image scanning interface and OmniPage for text scanning. The 300G is a flatbed scanner requiring a SCSI terminator when connected to the Mac. The fact that no terminator was supplied with the device was listed as a major shortcoming by our evaluator. Another shortcoming is the insufficient memory on the Mac for OmniPage to operate efficiently. (Although this isn't the scanner's fault, it is a consideration when you're putting the device to practical use--a minimum of 4MB is required).

Features noted as "best liked" include ease of use, low maintenance, better-than-average results for scanned graphics, and ability of the flatbed design to accommodate source documents with a variety of physical characteristics (e.g. books, charts, maps, etc.). Our study participant said he would recommend this configuration, with appropriate cautions with respect to memory and SCSI terminator requirements.

To overcome the problem of insufficient memory to process our test pages, the evaluator used a technique recommended by OmniPage. Text pages were saved as 300-dpi TIFF files (which, interestingly, all were 1 megabyte in size), then the ICR software was executed against those disk files. With this technique, the software 'reads' the text from disk, rather than having it passed directly from the scanner. The resultant test files were saved in MS Word format, which we subsequently converted to WordPerfect.

This material clearly demonstrated the suspect nature of manufacturers claims for text recognition accuracy. With an option turned on to record recognition accuracy during the scanning process, OmniPage reported 98-99.7% accuracy on several documents that were practically useless. As we discussed earlier in this report (third paragraph on page 6), these percentages represent the number of characters the software flagged as 'suspect,' but don't take into account those it incorrectly recognized. Nevertheless, several pages had few errors, either real or imagined. The Prestige Elite text and the Helvetica from a PC TAP Consumer Report page were particularly well done.

Summary

In conclusion, we'd like to add our own brief assessment of desktop scanning, gleaned through our experiences in this study. It appears there are a number of viable scanners on the market, and from what we've seen most of them do a reasonably good job at what they're designed for. After all, scanning technology has been around for a while, it just hasn't been in the desktop market until fairly recently. So you probably can find a low-end scanner that suits your needs for a list price in the \$2,000-\$4,000 range, and you can expect to get a reliable piece of equipment. However, the key to the utility of that piece of equipment is in the software you obtain to process the text or images the scanner can capture.

A number of good software products are available, each of which has its own capabilities and limitations. Many—but not all—scanners are sold with bundled image-processing software, and reasonably-priced products are available for those that aren't. With OCR products, though, the choices are wider and more varied. The better ones use intelligent character recognition techniques; these often come with a board that has software and additional memory where the ICR processing can be sped up without a lot of I/O to your computer. They have the power to deliver accurate text recognition at acceptable speeds, given your source documents are reasonably clear and sharp. These products presently list in the \$2,000-\$4,000 range. If your needs are more modest, there are some excellent performers for under \$1,000, but you must be prepared to accept their limitations in terms of text recognition and processing power.

This report has included a lot of descriptive text, and rather than concluding with more narrative we prepared a brief table. In deciding what to include in the table, we asked ourselves what a prospective scanner buyer would be asking him- or herself. These questions came to mind:

1. What type of scanner is it?
2. Will it work with my computer?
3. What is required to connect it to my computer?
4. Does any software come with it?
5. How much does it cost?

The table on the next page summarizes the answers to these five questions. If you want more details about a particular scanner or software product, refer back to the text in the body of the report.

Happy scanning!

**Desktop Scanners
Summary of Features**

Scanner	Type	Platform	Bundled Software	Available Interface	Price*
Apple	Flatbed	Macintosh	Image	SCSI	\$ 1,700
Chinon DS-3000	Portable, Overhead	PC	Image	1/2-board	\$ 995
Datacopy Model 830	Flatbed	Mac, PC	Image	SCSI, 1/2-board	\$ 2,900
DEST PC Scan	Sheetfed	Mac, PC	Text, Image	SCSI, 1/2-board	\$ 2,250
DEST PC Scan Plus	Sheetfed	Mac, PC	Text Image	SCSI, 1/2-board	\$ 2,500
DEST Model 202	Sheetfed	Stand-alone	Text-only Device	Serial Port	\$10,000
HP Scan-Jet Plus	Flatbed	Mac, PC	Text, Image	SCSI, Comm, Full board	\$ 2,000
Kurzweil 4000	Flatbed	Stand-alone	Text-only Device	Comm Interface	Not Avail.
Kurzweil 7320	Flatbed	Mac, PC	None	SCSI, Full board	\$ 4,995
Microtek MSF 300A	Flatbed	Mac, PC	None	SCSI, 1/2-board	\$ 3,000
Microtek MSF 300G	Flatbed	Mac, PC	None	SCSI, 1/2-board	\$ 3,495

*Figures are from available sources and may not reflect current market prices. They are included here only as a rough guideline to aid in product comparisons.

List of Study Contributors

Earl Beam
EPA National Enforcement Investigations Center
Denver Federal Center
Denver, CO 80225
(303) 236-5122 (FTS) 776-5122

Denise Cheatum
EPA National Enforcement Investigations Center
Denver Federal Center
Denver, CO 80225
(303) 236-5122 (FTS) 776-5122

Angela Edwards
Health Effects Research Laboratory
EPA Environmental Research Center
Research Triangle Park, NC 27711
(919) 541-4911 (FTS) 629-4911

Don Gorton
Information Center Consultant
EPA Region VIII
999 18th Street
Denver, CO 80202
(303) 293-7546 (FTS) 330-7546

Sophia Jeffries
UNC Graduate Assistant/IC Consultant
Information Centers Branch, MD-35
EPA National Computer Center
Research Triangle Park, NC 27711
(919) 541-3661 (FTS) 629-3661

David Levesque
Information Center Consultant
EPA Washington Information Center
401 M Street SW
Washington, DC 20460
(202) 475-7413 (FTS) 475-7413

Theresa Rhyne
Information Center Consultant
Information Centers Branch, MD-35
EPA National Computer Center
Research Triangle Park, NC 27711
(919) 541-0207 (FTS) 629-0207

List of Study Contributors

Robert Root
Information Center Consultant
EPA Washington Information Center
401 M Street SW
Washington, DC 20460
(202) 475-7413 (FTS) 475-7413

Diana Smith
Information Center Consultant
EPA Region IV
345 Courtland Street
Atlanta, GA 30365
(404) 347-0509 (FTS) 257-0509

David Taylor
PC TAP Coordinator
Enformation Centers Branch, MD-35
EPA Environmental Resarch Center
Research Triangle Park, NC 27711
(919) 541-0568 (FTS) 629-0568

Dr. Bellina Veronesi
Health Effects Research Laboratory, MD-74B
EPA Environmental Research Center
Research Triangle Park, NC 27711
(919) 541-2795 (FTS) 629-2795

Editor's note: Although *Open Line* doesn't appear in this Report, this information is included for those who may wish to contribute an item for a future Consumer Report.

How to Submit Items for Open Forum

In keeping with the PC Technology Assessment Program's objective to have the user community actively involved in TAP projects, users are encouraged to submit items for inclusion in future *PC TAP Consumer Reports*. If you have independently investigated the capabilities of a software product or a hardware component, we would like to hear from you. We'd also like you to share with others your solutions to any problems you may have encountered with a particular application or device, and about tricks, shortcuts, or unique applications you have devised. Although we can't promise to publish every contribution, we will evaluate them all in terms of their potential interest to our readers and their conformance to the spirit and intent of PC TAP.

There are no additional rules for *Open Forum* contributions, but here are some guidelines:

1. Contributions must be typed. Our first preference is that they be submitted on a floppy disk in WordPerfect format. If that isn't possible, the next best method is to EMAIL the text to DAVE.TAYLOR, EPA3099. The least preferable method, but still acceptable, is to mail a typewritten article to TAP at the address on the cover of this publication.
2. The length of your contribution will be determined somewhat by its complexity. However, keep in mind that we're primarily interested in the purpose of your study project and how pleased you were with the results, not in the nitty-gritty details of how you did it. We will publish your name, address, and phone number for those who want more details. Two to three pages is probably a reasonable maximum length. On the other hand, a paragraph containing a nugget that may be useful to others would be equally welcome.
3. All material submitted by users is subject to our editing, and you will not be given an opportunity to review the final manuscript before publication. Sorry, you'll just have to trust us. If we have questions or don't understand any part of your text, we'll contact you for clarification.

We hope you enjoy *PC TAP Consumer Reports*, and we look forward to hearing from individuals who have insights or discoveries to share with others. Thanks for your interest and your participation in the PC Technology Assessment Program.

