Advanced Character Recognition™

ACR™ Technology and Performance Analysis White Paper

For more information, please contact CharacTell:

USA and Far East
Dr. Paz Kahana

(617) 965-1014
pkahana@characTell.com

Europe and Israel
Mr. Ofer Comay

+972-3-648-9484
ofer@characTell.com

Table of Contents

List of Figures

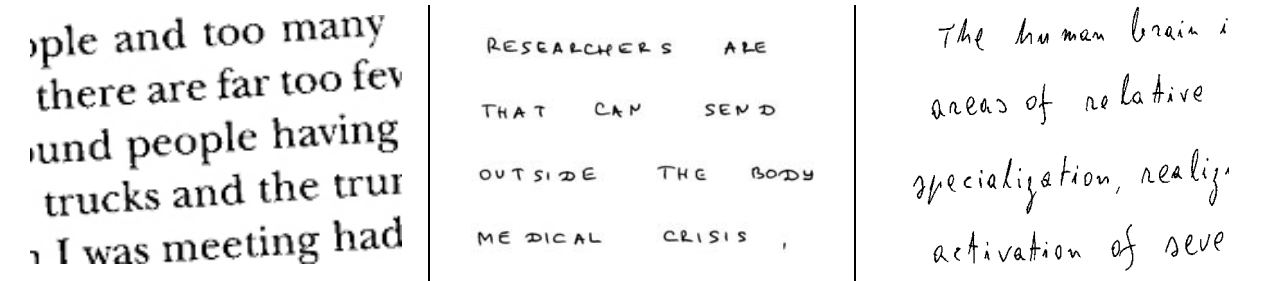
List of Tables

Introduction to Advanced Character Recognition

Automated conversion of characters found on paper to electronic (editable) form has existed for nearly five decades. Its first useful, commercial implementation was known as Optical Character Recognition, or OCR, and addressed recognizing machine-printed (books, output from high quality printers, etc., see Figure 2 below) materials. An early application was automated machines that read books for the blind.

As the technology matured and improved, its applications broadened and included widespread reduction, almost elimination, of retyping of existing documents, conversion of faxed documents, etc. Users have become accustomed to near perfect results (99.8% to 99.9% correct recognition results and retention of page attributes) when processing clean, machine-printed characters (Figure 1, left).

The next hurdle was the recognition of hand-printed characters (Figure 1, middle). A new term - Intelligent Character Recognition (ICR) - was coined in the eighties to distinguish this application from OCR. Due to the increased complexity of reading hand-printed characters, results are much more dependant on specific conditions such as image and writing quality, and context, and generally are lower than OCR's. ICR results ranging from 85% to 95% correct recognition (characters) are considered acceptable, if they provide a more economical model for the entry of the data involved in comparison to manual entry.



Optical Character Recognition (OCR) addresses high quality, machine printed characters	Intelligent Character Recognition (ICR) addresses hand-printed, upper-case characters	Advanced Character Recognition (ACR) addresses machine-printed, hand-printed AND hand-written, but non-connected characters

Figure 1 - The progression of character recognition from OCR to ACR

OCR and ICR have been the pillars of all practical recognition applications to date, and have been greatly successful. During the last two decades, it had been believed that OCR and ICR technologies have reached their maximum capabilities. Experts had not projected significant additional improvements in the automated conversion of scanned characters to ASCII format. Defying this belief, CharacTell has developed, and is now offering, what was previously considered impossible – both improved performance of traditional recognition applications, and new capabilities that can enable new applications.

CharacTell has developed a revolutionary approach to recognizing characters, commonly surpassing ICR accuracy, and expanding its applicability. Advanced Character Recognition™ (ACR™) allows reading of machine printed, hand printed and non-connected handwritten characters (Figure 1, right) with no supporting information in a variety of document types: notes, forms, letters and other mailed pieces.

When applied to traditional recognition applications such as form processing, ACR regularly delivers improved results over these obtained even by the best-of-class products of today. CharacTell has tested and verified this with respect to several leading commercially available products using a wide range of character sets. When applied to handwritten characters, ACR opens the way to new applications altogether such as converting pages full of notes to editable format, and can significantly improve performance in others such as man-machine interaction where handwriting can be used. ACR is the first technology to support this broad range of applications, and to offer recognition results high enough to improve on existing ones and enable commercial applications for reading of handwriting.

Advanced Character Recognition Technology

The recognition of lower case written letters has been considered unachievable. Human handwriting includes letters that may have ambiguous meanings. Many people write the letter "e" in exactly the same way that others write the letter "c." Some write the letter "r" close to the way others write the letters "v" or "c." CharacTell's Advanced Character Recognition (ACR) can effectively overcome these difficulties, breaking new grounds in this important area.

Traditional Optical Character Recognition Technologies

Traditional recognition follows a few basic steps, use common methods and expect certain conditions:

- **Character learning –** Every recognition system must have some a-priori knowledge about the characters it will find. Alphabets, numerals and special characters are pre-defined uniquely for all the font(s) and language(s) desired, and stored in the system's database. The result is that large samples of characters are needed for learning to allow good performance. The "learning curve" (no pun intended...) is directly dependent on the size of the learning set, often tens of thousands.
- **Feature extraction –** The bitmap of each character (segmentation is assumed) is analyzed for features that define it. Because of the varying nature of character writing and image quality, it is best to search for invariant features that uniquely define the character.
- The exact set of features sought for each character is determined by the recognition system's architecture and algorithms. Typical examples are the number of peaks (maxima) or bottoms (minima), line slopes, the number of intersections between the character and various reference lines, etc. It is common to need to extract hundreds, or even thousands of features from each character.
- **Character classification –** Once the features are extracted, they are investigated using many methods, such as neural network. Each character is compared to known shapes in the system's database. This process is referred to as classification. The classifiers are part of the core technology of each recognition system and, along with the feature extraction process, define their capabilities and limitations.
- **Contextual analysis –** Rarely are characters used in isolation. Since in the majority of applications additional information is available, it can be used to help improve recognition results. For words, dictionaries or lexicons can eliminate a character error, and still provide an accurate recognition of the word. Postal address syntax requires that state, city, and zip code information all be consistent. And in forms, there is often a set of logical rules that links one or more fields. Additional information of this nature provides crucial guidance and improves results.
- **User verification –** Despite all the advances in technology, there will always be instances when no automated recognition process can successfully convert a written character. The handwritings of these authors are prime examples, where even the human mind often fails to understand their intentions... Data then must be presented for inspection and corrected by a person operating with additional information and applying best human judgment. This is critical; especially in applications - such as financial - where 100% accuracy is always required.

At the conclusion of the sequence of these steps data in inaccessible image format has successfully been converted to useful format and can be further processed. The recognition is deemed complete.

A Closer Look Under the ACR Hood

Recognizing a character is the equivalent of knowing the family it belongs to. Determining that a particular character is an "a" or a "Y" or a "9" means that it belongs to the family of shapes corresponding to that letter or number. Traditional technologies employ techniques known as feature extraction to create a unique definition of each character, and neural networks to match features extracted from characters to known features of characters the engine expects to find.

Not long ago, if we wanted to determine which family a person belongs to, the only practical way to do so was to extract that person's characteristics (eye color, bone sizes, height, facial shapes, etc.) and compare them to these of as many people of each possible family as possible. The larger the number of features extracted and compared, the higher the accuracy of the recognition can be. Today, however, there is a better way. We sample the person's DNA and compare it to the DNA of a few people (or even one only) from each family.

By applying similar techniques to character recognition, CharacTell invented a new way of obtaining a chain of symbols from each character that can be used as "DNA" with substructure, much like "Genes." The underlying premise is that each character – despite the widely varied ways it can be represented – contains one or more "Genes" that uniquely identify it from a total number of twenty-eight possible "Genes." Finding and identifying as few as four or five such "Genes" can lead to high confidence of successful recognition.

Figure 2 – Different shapes of the same character contain the same "Genes"

During the learning stage, a database of the "Genes" of each character is created. Each Gene is given a certain weight for all possible recognition results. ACR then creates a string of symbols from each character it recognizes. These strings (of Genes) are compared to the known (DNA) structure. *This allows more accurate recognition confidence factors, offers better results, and requires much smaller samples for training*.

ACR is truly a breakthrough technology. By departing from established methods, CharacTell has been able to create a core technological foundation strong enough to be the basis for a broad next generation of proprietary recognition products. The algorithms and techniques used by ACR are strict trade secrets of CharacTell, but their ramifications in terms of user benefits, competitive advantages, and differentiation from all other products on the market today are tremendous:

- **Learning of new languages and characters** – ACR has a unique ability to train itself quickly using an extremely small sample of characters. Ordinary recognition engines that rely on feature extraction techniques routinely need as much as 50,000 samples (or more) per each character they learn. Vendors boast that they use sample sets consisting of millions of characters. The collection of such a huge number of real samples is costly and time consuming to vendors, and prohibitive to users, often prevents entry to geographical markets where additional language support is needed. ACR needs only a few thousand characters to fully learn a new language, a new font, or special characters of a language (presently, of Latin-based languages). This reduces the cost of use to vendors and offers users increased flexibility in how they can apply ACR-based products.
- **Learning of individual writing styles** – Most every person's handwriting is unique. ACR accurately learns the individual handwriting style of each user, leading to very high recognition rates where existing technologies fail. This is a strategically important capability when applied to the man-machine interaction problem, as users will see a substantially increased recognition of their writings.

- **Continuous learning –** Where applicable and possible, ACR can continue to learn the characters it recognizes beyond the initial training. The learning process continues as long as the system is in operation, as it collects comparisons of questionable characters with their correct identity. The result is continued improvements in performance while as the system is being used.
- **Very high recognition and lower error rates –** Characters may be uniquely identified based on the presence of only a small number of identifying "Genes" in the string of symbols that represent them. This leads to correct recognition results by the engine, and a high confidence level that it correctly recognized the character. Error rates (characters wrongly recognized and not flagged as questions) are greatly reduced, offering substantial savings during the data validation and correction phases.
- **Unique contextual analysis –** With CharacTell unique method of using commonly available knowledge, such as standard language dictionaries or databases, substantially enhanced recognition can be achieved, even when individual character conditions would not have supported it.
- **Advanced voting –** ACR also includes a sophisticated, new voting scheme that allows additional engines to collaborate when needed. Whenever possible it is recommended to use more than one engine and vote among them for best results. This means that users can easily apply ACR in conjunction with existing recognition products and realize significant performance improvements, while making minimal changes in the architecture of their products and/or systems.
- **Grayscale processing –** Critical for characters made with a blue pen, ACR utilizes special algorithms that work on grayscale images, and then convert them smartly into binary format. ACR's proprietary algorithms analyze pad lines, words and connected characters.

In total, ACR's claim to fame are its breakthrough performance in recognizing handwritten characters, the improvements it delivers in interpreting hand printed characters over today's existing products, and its unique ability to learn new languages, special characters and individual writing styles. ACR excels where other products offer just OK performance – a level that has become acceptable because no one thought it could be improved - and it particularly extends recognition reach in the area of reading handwritten text without additional context information. *This makes ACR the most dynamic and broadly applicable recognition technology available today*.

CharacTell Products Performance Comparison

CharacTell Products

At this time, CharacTell markets numerous products, all based on ACR technology[1]:

- **JustICR –** Targeting the traditional form processing market, JustICR was CharacTell's first product. It is a library of programmable functions offered as an open API that allow developers of form processing applications to quickly and easily integrate advanced recognition capabilities. JustICR has been widely used by CharacTell's customers. JustICR has been installed successfully at many household name organizations in the United States, Switzerland, Germany, Brazil, India, Italy, Ireland, Kenya, Cyprus, Israel, Portugal, and Croatia, to mention just a few. In those, and other installations, JustICR performs difficult recognition tasks that are not possible under ordinary conditions, or by non-ACR engines.
- **iREAD Forms –** Introduced to the form processing market in March 2002, iREAD Forms is a natural evolution of JustICR. In addition to including CharacTell's core technology, iREAD Forms integrates an additional OCR/ICR engine, which works in conjunction with the CharacTell engine. The results generated by each engine are combined for even superior and more accurate overall results. The combined engine offers improved recognition in terms of higher correct recognition percentages and substantially reduced errors. This further demonstrates CharacTell's ability to offer superior, accurate, and easy-to-use recognition products that are highly competitive in the marketplace.
- **SoftWriting –** SoftWriting is an end-user application program that allows scanning in, or importing of, pages containing hand-written notes, and converting them into RTF format that is editable by most popular of word processing products (such as Microsoft Word, Lotus WordPro, Corel WordPerfect, etc). SoftWriting was nominated for the Product of the Year Award for 2001 by PC Magazin (Germany) and has been favorably reviewed in the media for its advanced recognition capabilities and ease of use.
- **FormStorm Enterprise** - Based on iREAD Forms, FormStorm Enterprise is a robust, flexible, and function-rich application software designed for automated form processing and data entry. FormStorm Enterprise is designed with the goal of great ease of use, while taking full advantage of the power afforded by ACR and iREAD Forms. To address specific areas of functionality, additional engines and technologies have been added.

CharacTell Products Performance

Table 1 provides a summary of the results achieved by applying the relevant CharacTell products to the character samples shown in the Appendix. Page 1 contains representative upper case, lower case and numeral characters one finds in various applications where characters are in isolated strings, such as in forms. Pages 2 through 4 contain three examples of notes written in lower case handwriting (pages 3 and 4 are included as images in the SoftWriting v3.2 evaluation program available from CharacTell. Page 1 was subjected to reading by JustICR, and pages 2 through 4 to SoftWriting version 3.2.

[1] ACR, Advanced Character Recognition, SoftWriting, iREAD Forms, FormStorm, and CharacTell are trademarks of CharacTell Ltd. 1998 – 2003.

Document	Words in original documents	Words in output document	Correct	Incorrect	Added	Total Incorrect	% Correct	Characters	Characters in output document	Correct	Incorrect	Added	Total Incorrect	% Correct	Comment
	Words							Characters							
Page 1a	Not applicable							1202	1202	1182	20	0	20	98.34%	Lower case
Page 1a	Not applicable							977	977	964	13	0	13	98.67%	Upper case
Page 1a	Not applicable							1871	1871	1838	27	-6	33	98.24%	Numerals
Page 2a	94	95	92	2	1	3	96.81%	562	563	548	9	1	10	98.22%	Lower case
Page 3a	141	141	138	3	0	3	97.87%	660	659	650	17	-1	18	97.27%	Lower case
Page 4a	89	90	1	2	1	3	96.63%	413	411	405	10	-2	12	97.09%	Lower case
Totals	324	326	231	7	2	9	97.22%	5685	5683	5587	96	-8	106	98.14%	

Table 1 – Performance results of CharacTell products on the sample pages in the Appendix

SoftWriting is capable of converting between 97% and nearly 99% of all characters in what is obviously diverse and difficult to read handwritten material without the aid of any context information other than a standard language dictionary. On a word basis, the results are also high, ranging around 97%.

That means that on the average, only six (6) words[2] in a page containing two hundred (200) words will need any kind of correction by the user. Such a level of performance appears to justify the application of software to the automated conversion of this type of characters.

CharacTell Form Processing Products Performance

CharacTell has tested the performance of its high-end form processing product, iREAD Forms, and compared it to results obtained from two leading competitive products. These products are identified as Engine A and Engine B. Table 2 shows test results and comparisons made between February 2002 and April 2002.

Application	Engine A			Engine B			iREAD Forms			IRF-A	IRF-B
	rec	err	FOM	rec	err	FOM	rec	err	FOM	FOM	FOM
Germany handprint numerals	97.11%	1.44%	89.90%	98.76%	0.82%	94.64%	98.87%	0.72%	95.26%	**5.36%**	**0.62%**
Germany handprint uppercase	84.54%	5.42%	57.42%	88.13%	3.58%	70.21%	90.69%	3.58%	72.77%	**15.35%**	**2.56%**
Germany handprint upper and lowe	77.45%	10.98%	22.55%	87.02%	5.16%	61.23%	87.35%	4.49%	64.89%	**42.35%**	**3.66%**
Germany handprint upp/low names	89.85%	5.24%	63.64%	95.59%	1.66%	87.27%	98.25%	1.16%	92.43%	**28.79%**	**5.16%**
US handprint uppercase	87.00%	2.90%	72.52%	86.75%	4.22%	65.65%	95.20%	2.24%	84.02%	**11.51%**	**18.38%**
US handprint numerals	96.21%	0.66%	92.89%	98.20%	0.66%	94.88%	98.20%	0.66%	94.88%	**1.99%**	**0.00%**
Portugal numerals (low quality)	91.68%	4.16%	70.87%	91.08%	1.43%	83.95%	91.56%	1.43%	84.42%	**13.56%**	**0.48%**
Portugal handprint uppercase	85.52%	5.29%	59.08%	84.83%	8.74%	41.15%	92.18%	3.45%	74.94%	**15.86%**	**33.79%**
Israel numerals	93.78%	1.33%	87.13%	93.34%	1.44%	86.13%	97.23%	0.55%	94.45%	**7.33%**	**8.32%**
One font, medium quality, numerals	100.00%	0.00%	100.00%	97.99%	0.00%	97.99%	99.87%	0.00%	99.87%	**-0.13%**	**1.87%**
OCRA, numerals, low quality	95.00%	2.00%	85.00%	97.25%	0.00%	97.25%	99.00%	0.00%	99.00%	**14.00%**	**1.75%**

Table 2 - Performance comparison between CharacTell's and leading competitive products

Each line in Table 2 corresponds to a test suite consisting of a large number of characters of different nature from different geographical source and thus presenting a different writing style. For each suite, the performance of each engine was first tested separately and compared. Subsequently, the performance of the integrated pair was compared to the performance of the best of the two engines. Results were consistent: *In the vast majority of tests, the CharacTell engine outperformed the competitive engine individually.*

In nearly all the tests, the integrated pair consisting of the CharacTell engine and the competitive one, performed better than the best performing single engine for that test. The overwhelming conclusion of these tests is that CharacTell's products offer state-of-the-art performance, and when integrated along with additional engines, performance rises to levels not previously achieved.

[2] Exact results depend on a number of conditions, including image quality, character segmentation and separation, the level of user training performed and adherence to proper operation of the software.

Comments to performance comparisons

Full Page Text Tests (Pages 1 – 4)

The character samples shown in page 1 constitute a small portion of a data set containing thousands of characters. Recognition of characters other than numerals was aided only by a standard language dictionary.

1. No edits to ACR output were made, except pre-defining abbreviations: (vel = velocity, aber = aberration) on page 3.
2. Incorrect counts include rejects (low positive confidence) and false positive results.
3. Spaces between words are not included in calculation (including spaces would have had a positive effect on calculated values).
4. Added or subtracted words and characters are considered as errors in their respective categories.
5. Calculation basis is the number of original words and characters for their respective categories.

Form Processing Tests

1. Eleven different tests were performed using sets of different types of well-segmented characters.
2. Because of the nature of the tests and the definition of the parameters involved, the results are most meaningful in the context of form processing applications where designated fields contain somewhat defined data types.
3. Performance results for each separate engine, as well as the combined engine, are expressed in terms of measured percentages of correct recognition (rec) and error rates (err), and a calculated figure of merit (FOM) for each of the eleven tests.
4. FOM = [% correct recognition] – 5 * [% error rate]. The FOM combines results of correct and incorrect recognition into a single number that penalizes more heavily (by a factor of 5) for recognition errors. This is due to the higher costs of identifying and correcting errors after the fact.

The right-most column shows the difference in FOM – what one might think of as the performance improvement – between the best performing engine of the two separate ones and a combined engine consisting of JustICR and each one of the tested engines.

No other context information other than standard language and name dictionaries was used. When dictionaries were used, they were applied to both engines natively to the respective products.

Appendix 1 – CharacTell in World News

Seattle Times

Source: http://seattletimes.nwsource.com/html/businesstechnology/134336164_pthandwrite02.html

Sunday, September 02, 2001 - 12:00 a.m. Pacific

Reviews

Remarkable new program scans your handwriting

By Craig Crossman
Knight Ridder Newspapers

Ever since the dawn of Apple's now-defunct Newton personal digital assistant, a machine able to transcribe handwriting into typewritten text has remained pretty much a dream.

Today's popular PDA's like the Palm use a shorthand form of handwriting called Graffiti. It takes a while to learn because some letters and punctuation must be formed differently. The Palm also requires you to enter each letter in the same surface area rather than alongside of each other.

So instead of the natural movement from left to right as you write, your hand remains stationary as you overwrite each character. And of course, you must carry the PDA wherever you plan to use it.

But now there's another way.

SoftWriting is a remarkable new program that reads handwriting in much the same fashion as optical-character-recognition (OCR) software scans typewritten text. With SoftWriting, you simply place your handwritten document into a scanner and watch your words appear typed into your word processor. You can even stylize your writing. For example, words you underline will appear as underlined typewritten text. Or you can instruct SoftWriting to apply most any style such as boldface or outline to anything you underline.

If you've drawn something such as arrows, a picture or anything else on your document that you don't want to be transcribed, you can tell SoftWriting to keep it as a graphical image and move it into the final electronic document just as it appears on the scanned paper.

SoftWriting requires no training out of the box. As you continue to use the program, it continues to learn your handwriting and accuracy improves. All of this occurs in the background without requiring your attention. Claimed accuracy is around 98 percent.

One thing you will have to learn is that each character you write must be disconnected from the others. You don't have to print although printing will work as well.

Downloadable from the Charactell Web site, www.charactell.com, SoftWriting is fully functional for 14 days so you can try before you buy. If you like it, $59.95 buys the unlocking code. It requires Windows.

The Review Zone's Best of 2001

Source: http://www.thereviewzone.com/charactertellsoftwriting3.html

Teach your computer how to read your unique handwriting and you'll never type a term paper again!

You're in the library when suddenly it hits you. Something clicks (and it's not your computer mouse!). The term paper you've been procrastinating is starting to shape up and you're minus a computer and floppy. You start writing sentences and before you know it, you're writing paragraphs and citing references. Then, the realization hits you. You'll have to go home and type the paper on your computer. What a waste of time, considering the paper is due the next morning!

Enter CharacTell's SoftWriting 3.0. If you're a PC user and you'd like to convert your handwritten notes from paper into a Microsoft Word doc or as a text document, you're in luck!

Be sure that your letters are non-connected letters (upper and lower case block letters) and "train" the software to recognize your writing! Whether you'd like your lab notes typed, you're stuck in a lecture hall taking notes without your notebook computer, or you're taking minutes in a boring Homeowner's Association meeting, this software is intelligent! It can "learn" your handwriting style and your vocabulary, as well.

The neat thing about this program is that you can teach the software to pick up abbreviations (instead of writing for example, you can write ex). The Intelligent Character Recognition (ICR) rate is 98 percent per character (after inputting documents with approximately 500 words). The system keeps learning your writing as you scan documents. Sketches and diagrams can also be transferred into the document you're working on.

Obviously, you will have to edit the text before you turn in your winning paper! The software helps you identify potential errors (like a Spell-Check program) and provides you with other word choices. If you aren't working on completing a paper, this program is ideal to regularly enter your own notes in English, German, Spanish, Portuguese, and Dutch.

I thought that this program would be expensive! If you're a student, you can get a CD-ROM for less than $30 (otherwise, you'll have to fork over $60). Considering that developing ICR programs requires sophisticated programming, including "highly adapted algorithms", it's quite a bargain. The biggest challenge for me was writing in a non-connected manner. When I'm in a hurry, I tend to connect letters in cursive. To use this program, your handwriting MUST be non-cursive.

Don't miss the four-minute "demo" on the CD-ROM before you begin to use this software. You can also download a 14-day free trial at the company's web site, http://www.charactell.com

At press time, this product was available for PC users only. If you're a Mac user, keep typing!

Multimedia-test.de

Source: http://www.multimedia-test.de/1024/software/office/softwriting.htm

English translation: http://www.charactell.com/temp/20010701.htm

SoftWriting

by Martin Klinkigt, www.multimedia-test.de

Why not to let your computer do the typing

With SoftWriting you can have your handwritten texts transformed, so that you can, for example, process these texts in Word. All you need is a computer and a scanner.

Transformation

The transformation takes place very quickly and easily. The papers are scanned by pressing a button and afterwards it is transformed to text. Now follows the correction-reading, where SoftWriting marks the words about which it is insecure. These have to be corrected almost always. If words are written in blue, SoftWriting is not 100% sure, although, here in most cases it is correct or only one letter is wrong. After correction the text is saved without any effort as a word-file and is ready for further work.

Exactness

The recognition rate is quite remarkable. One has to keep in mind, that this is no regular OCR text-recognition software but that the individual handwriting should be recognized. With a little bit of training of the software, after four scanned pages, a recognition rate of 99% is possible. When we tried it, the rate after two pages was already 70%.

Handwriting

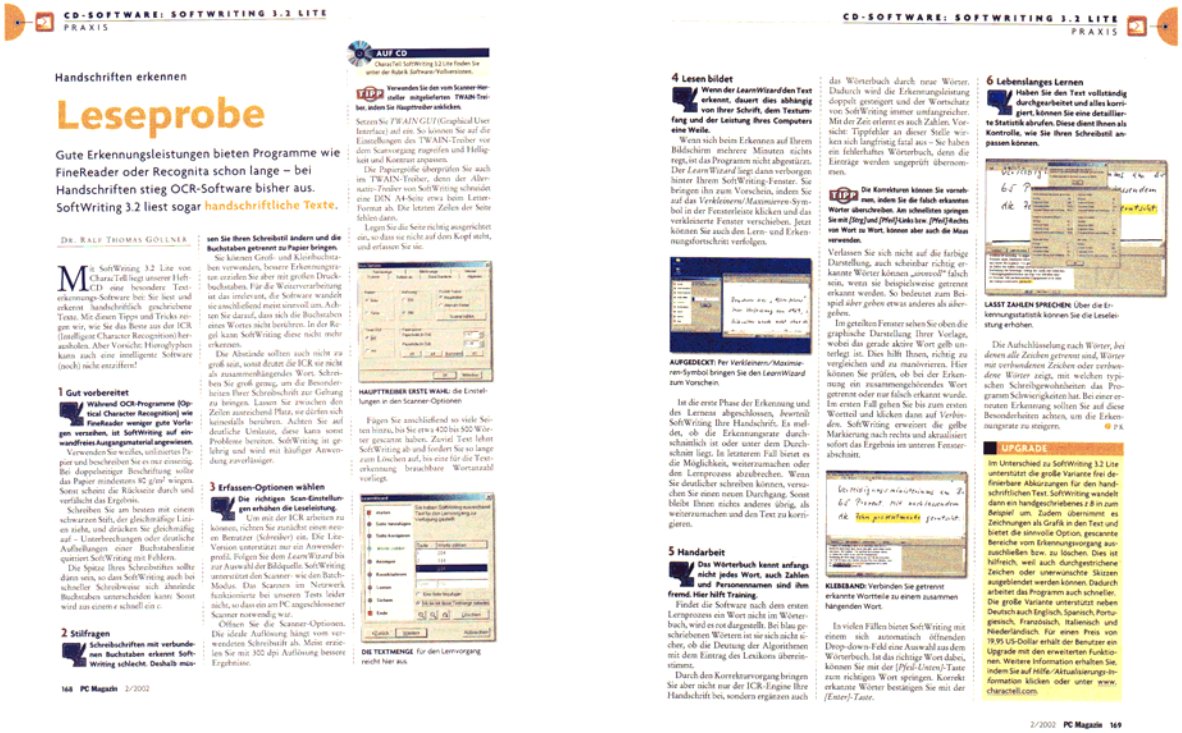
SoftWriting is concerning your handwriting very easy to satisfy. But with real script the software does not work. The first document is sent to the internet server without real benefit. Probably the recognition is too complicated. With handwritten block letters SoftWriting adjust itself very quickly to your handwriting. This way good results can be reached.

Conclusion

SoftWriting can be recommended to everyone who need their handwritten papers as text-files in the computer. If you want to be sure about your handwriting, CharacTell offers at its homepage a full functioning 14-day test version for download.

PC Magazin (PC Magazine, Germany) Product of the Year 2001 List

PC Magazin reviewed SoftWriting in February 2002, and placed it on the Product of the Year 2001 list:



CD-SOFTWARE: SOFTWRITING 3.2 LITE
PRAXIS

Handschriften erkennen

Leseprobe

Gute Erkennungsleistungen bieten Programme wie FineReader oder Recognita schon lange – bei Handschriften stieg OCR-Software bisher aus. SoftWriting 3.2 liest sogar handschriftliche Texte.

DR. RALF THOMAS GÖLLNER

Mit SoftWriting 3.2 Lite von CharacTell liegt unserer Heft-CD eine besondere Texterkennungs-Software bei: Sie liest und erkennt handschriftlich geschriebene Texte. Mit diesen Tipps und Tricks zeigen wir, wie Sie das Beste aus der ICR (Intelligent Character Recognition) herausholen. Aber Vorsicht: Hieroglyphen kann auch eine intelligente Software (noch) nicht entziffern!

1 Gut vorbereitet

2 Stilfragen

3 Erfassen-Optionen wählen

AUF CD

HAUPTTREIBER ERSTE WAHL: die Einstellungen in den Scanner-Optionen

DIE TEXTMENGE für den Lernvorgang reicht hier aus.

168 PC Magazin 2/2002

CD-SOFTWARE: SOFTWRITING 3.2 LITE
PRAXIS

4 Lesen bildet

AUFGEDECKT: Per Verkleinern/Maximieren-Symbol bringen Sie den LearnWizard zum Vorschein.

5 Handarbeit

KLEBEBAND: Verbinden Sie getrennt erkannte Wortteile zu einem zusammen hängenden Wort.

6 Lebenslanges Lernen

LASST ZAHLEN SPRECHEN: Über die Erkennungsstatistik können Sie die Lesele-istung erhöhen.

UPGRADE

2/2002 PC Magazin 169

The review said: “Other programs such as FineReader and Recognita have offered good performance for quite some time now, but with SoftWriting OCR software now rises higher. SoftWriting version 3.2 reads even handwritten texts.”

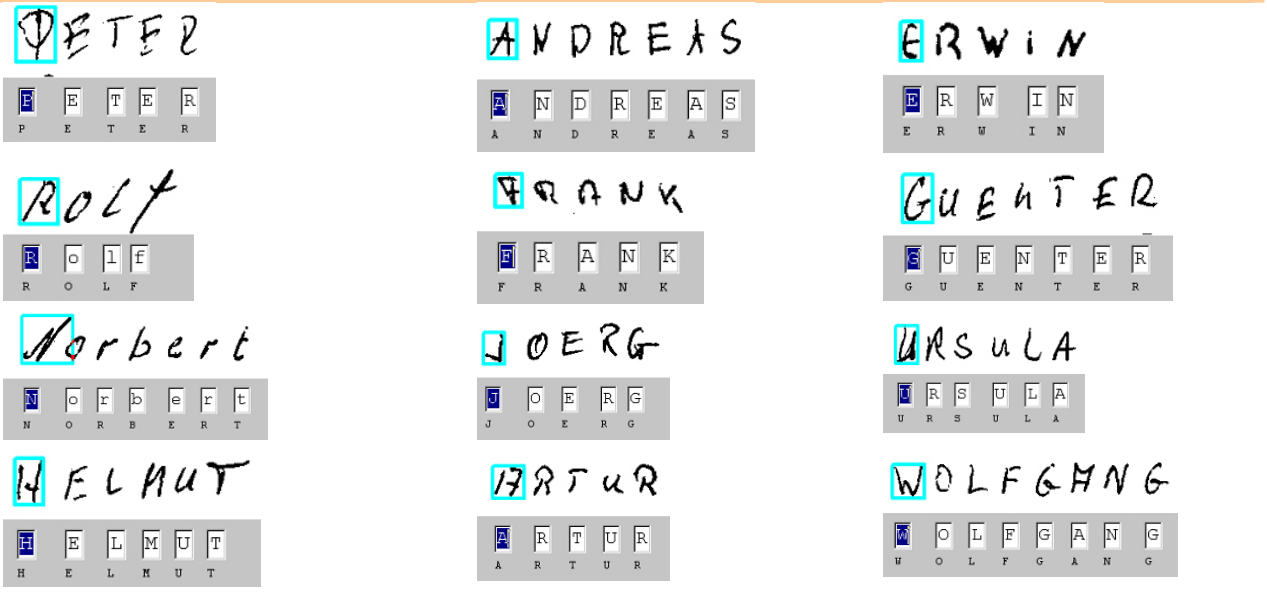
IMERGE Consulting, Arthur Gingrande

Arthur Gingrande (http://www.imergeconsult.com/resume/arthurgingrande.html) a well-known ICR consultant at Imerge Consulting, reports that, "*as an industry analyst specializing in intelligent character recognition, I've investigated a lot of claims about recognition accuracy of hand print characters in unconstrained environments - and SoftWriting really works! This is the first time I've seen an ICR application designed for ordinary, everyday use that performed so successfully on real world, handwritten documents.*" (http://www.charactell.com/temp/GingrandeWhitePaper.pdf).http://www.charactell.com/temp/GingrandeWhitePaper.pdf

Appendix 2 - Upper & Lower Case, and Numeral Character Examples

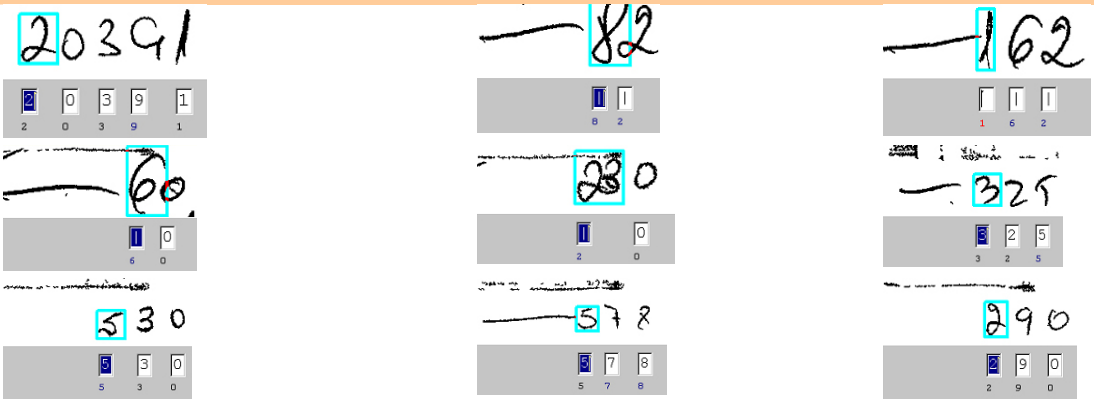
Page 1a

Upper case



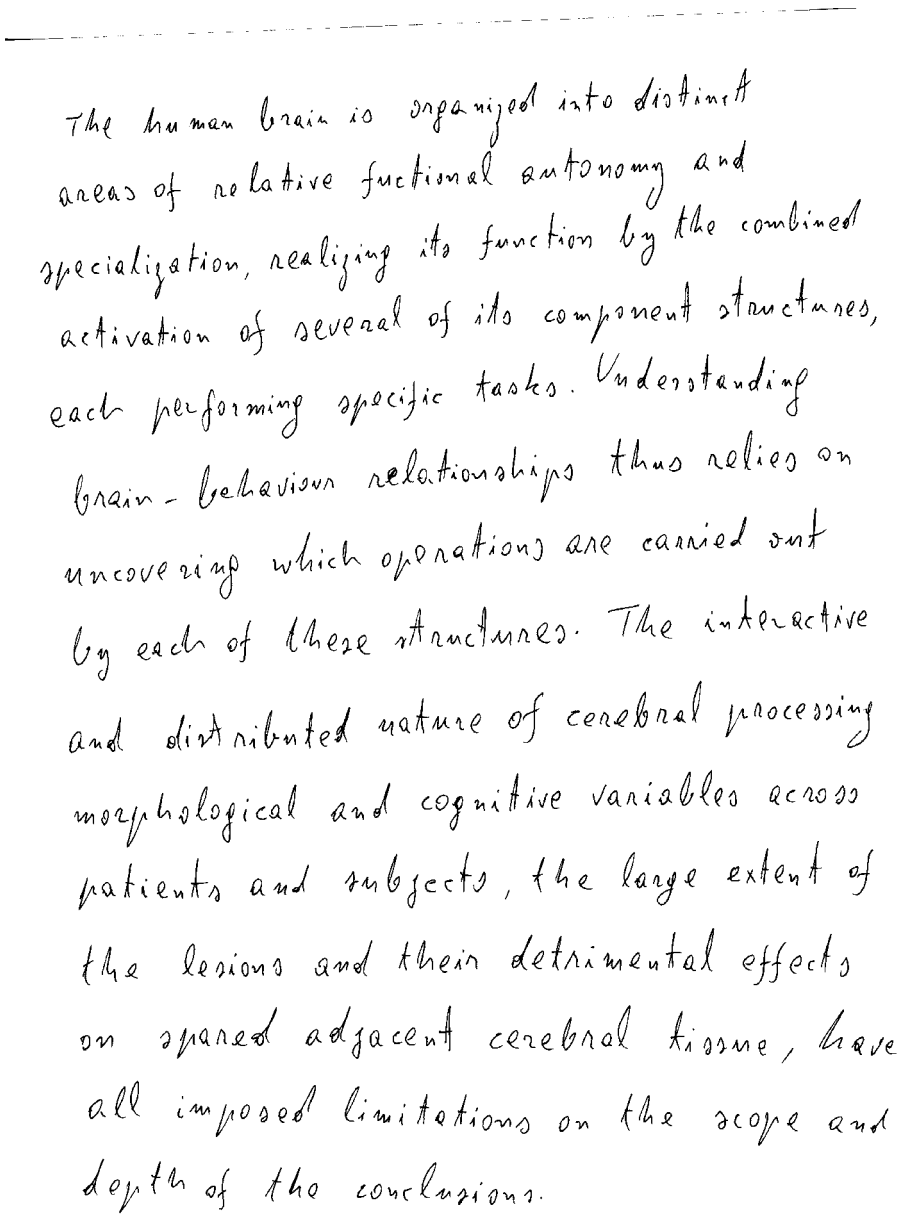
Lower Case

Numerals



Page 2a

Original text



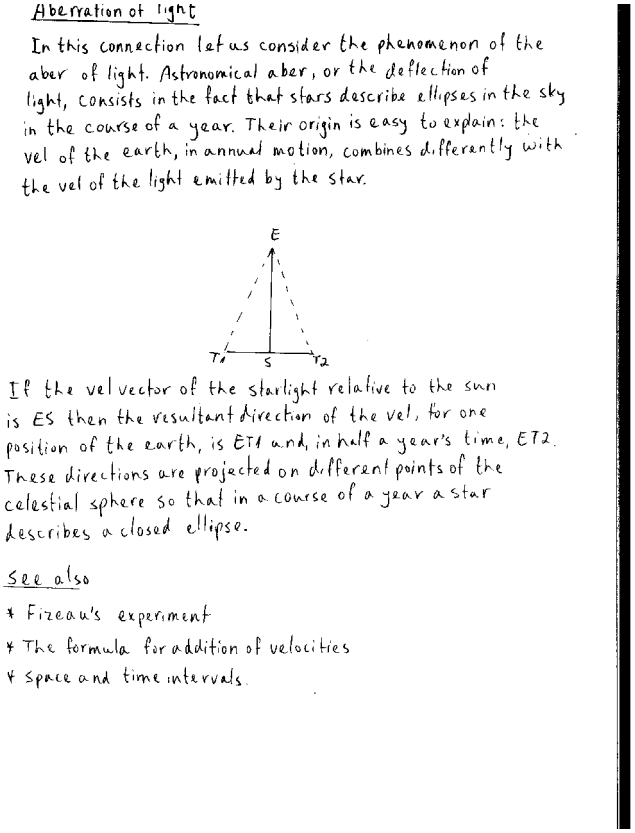
The human brain is organized into distinct areas of relative fuctional autonomy and specialization, realizing its function by the combined activation of several of its component structures, each performing specific tasks. Understanding brain - behaviour relationships thus relies on uncovering which operations are carried out by each of these structures. The interactive and distributed nature of cerebral processing morphological and cognitive variables across patients and subjects, the large extent of the lesions and their detrimental effects on spared adjacent cerebral tissue, have all imposed limitations on the scope and depth of the conclusions.

Recognition output results

The human brain is organized into distinct areas of relative fictional autonomy and specialization, realizing its function by the combined activation of several of its component structures, each performing specific tasks. Understanding brain - behavior relationships thus relies on anyone ring which operations are carried out by each of these structures. The interactive and distributed nature of cerebral processing morphological and cognitive variables across patients and subjects, the large extent of the lesions and their detrimental effects on spared adjacent cerebral tissue, have all imposed and stations on the scope and depth of the conclusions.

Page 3a

Original text



Aberration of light

In this connection let us consider the phenomenon of the aber of light. Astronomical aber, or the deflection of light, consists in the fact that stars describe ellipses in the sky in the course of a year. Their origin is easy to explain: the vel of the earth, in annual motion, combines differently with the vel of the light emitted by the star.

If the vel vector of the starlight relative to the sun is ES then the resultant direction of the vel, for one position of the earth, is ET1 and, in half a year's time, ET2. These directions are projected on different points of the celestial sphere so that in a course of a year a star describes a closed ellipse.

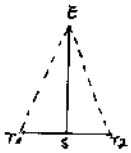
See also

* Fizeau's experiment
* The formula for addition of velocities
* space and time intervals.

Recognition output results

Aberration of line

in this connection let us consider the phenomenon of the aberration of light. Astronomical aberration, or the deflection of light, consists in the fact that stars describe ellipses in the sky in the course of a year. Their origin is easy to explains the velocity of the earth, in annual motion, combines differently with the velocity of the light emitted by the star.



!I the velocity vector of the starlight relative he the sun is ES then the resultant direction of the velocity, for one position of the earth, is ET1 and, in half a year's time, ET2 these directions are projected on different points of the celestial sphere so that in a course of a year a star describes a closed ellipse.

See also

· Someone experiment

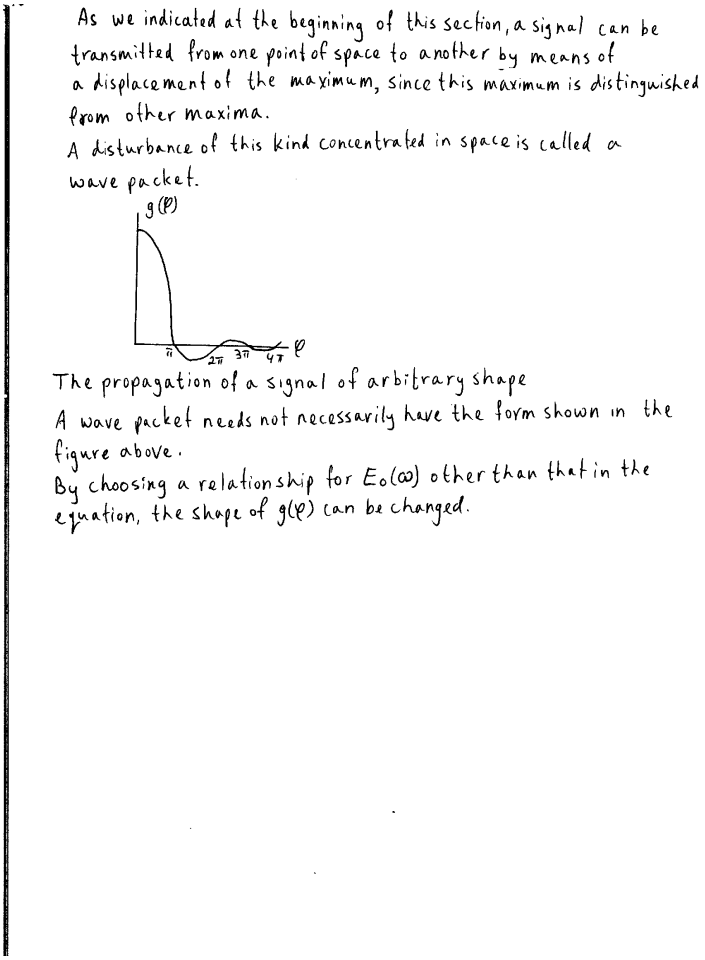
· The formula for addition of velocities

· Space and time intervals.

No edits to ACR output were made, except the pre-definition of abbreviations: (vel = velocity, aber = aberration).

Page 4a

Original text



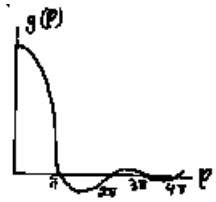
As we indicated at the beginning of this section, a signal can be transmitted from one point of space to another by means of a displacement of the maximum, since this maximum is distinguished from other maxima.
A disturbance of this kind concentrated in space is called a wave packet.

The propagation of a signal of arbitrary shape
A wave packet needs not necessarily have the form shown in the figure above.
By choosing a relationship for $E_0(\omega)$ other than that in the equation, the shape of $g(p)$ can be changed.

Recognition output results

As we indicated at the beginning of this section, a signal can be transmitted from one point of space to another by means of a displacement of the maximum, since this minimum is distinguished from other maxima.

A disturbance of this kind concentrated in some is called a wave packet.



The propagation of a signal of arbitrary shape a wave packet needs not necessarily have the form shown in the figure above.

By choosing a relation sup for $E_0(\omega)$ other than that in the equation, the shape of $g(p)$ can be changed.