International Journal on Document Analysis and Recognition
Schomaker, L. and Segers, E. (1999). Vol 2, pp. 13-18.

Finding features used in the human reading of cursive handwriting

Lambert Schomaker and Eliane Segers

Nijmegen Institute for Cognition and Information
P.O. Box 9104, 6500 HE Nijmegen, The Netherlands
schomaker@nici.kun.nl / http://hwr.nici.kun.nl

The date of receipt and acceptance will be inserted by the editor

Abstract. This paper first summarizes a number of findings in human reading of handwriting. A method is proposed to uncover more detailed information about geometrical features which human readers use in the reading of Western script. The results of an earlier experiment on the use of ascender/descender features were used for a second experiment aimed at more detailed features within words. A convenient experimental setup was developed, based on image enhancement by local mouse clicks under time pressure. The readers had to develop a cost-effective strategy to identify the letters in the word. Results revealed a left-to-right strategy in time, however, with extra attention to the initial, leftmost parts and the final, rightmost parts of words in a range of word lengths. The results confirm high hit rates on ascenders, descenders, crossings and points of high curvature in the handwriting pattern.

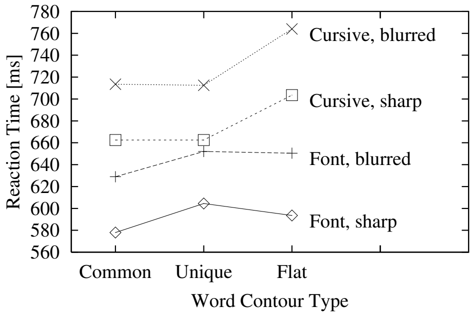
Key words: Human Reading – Perception – Features – Cursive Handwriting

1 Introduction

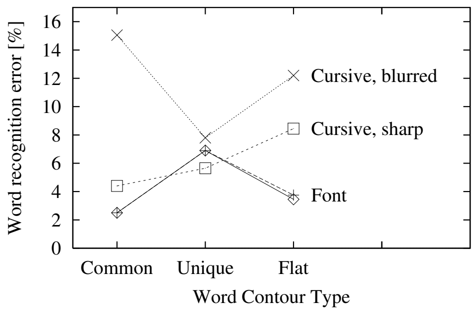
In the development of automatic recognizers of cursive handwriting it is useful to know what geometric features of script humans use in the reading process. However, the vast majority of studies in human reading is directed at the reading of machine-print characters. The human reading of machine-print letters is usually not influenced in a clear way [1] by the presence of ascenders and descenders[1] as is often assumed. The strongest argument comes from so-called case-mixing experiments [2] where human readers are presented with alternating patterns of upper-case and lower-case letters in words. This manipulation slows down the reading process without a clear deteriorating effect on reading accuracy. Most of these experiments are aimed at elucidating the question whether holistic word-based features are used in human reading, as opposed to an early segmentation into letters and the subsequent use of intra-letter features for classification. However, early research already indicates that there may be fundamental differences between the reading of machine print and handwriting [3]. Most psycholinguist models of the reading of machine-printed words assume a form of parallel processing, in which the features of individual letters are simultaneously used to activate words in a lexicon [4]. It should be noted that in machine print, the individual letters are easily segmented from the background, and all letter features are sufficiently clear such that the presence of an ascender or descender may not provide an inherently outstanding source of information as compared to other salient features present in a letter (closed areas, sharp endings, crossings). The human reading process of cursive script may be different in this respect. Here, the presence of ascenders and descenders **does** appear to enhance human recognition of isolated handwritten words [5]. Figure 1 shows the effects of word contour features on the human reading speed (reaction time) and error rate. The word contours were either (1) flat, or (2) consisted of a common pattern of ascenders and descenders in a large and representative text corpus, or (3) consisted of a unique (low-frequent) pattern of ascenders and descenders. Words (n=240) were presented to the readers (N=32) on a CRT screen using a machine font or cursive handwritten version, and were rendered in a sharp or Gaussian-blurred version.

Figure 1a shows that the reaction time goes up (i.e., reading speed goes down) in cursive-written words which have no ascenders or descenders (condition 'flat') but the reaction time is hardly affected by the word contour in machine-printed words. Blurring has a general decelerating effect on reading speed for both script types. As regards recognition accuracy (Figure 1b), it can be observed that sharply rendered cursive words with a flat contour are recognized less accurately than a machine-print font. When the word image is blurred, the cursive words with common or flat contours suffer most from

[1] Ascenders are defined as letters with large vertical strokes extending well above the corpus, or 'x' size. Descenders are defined as letters containing large vertical strokes extending well below the base line of handwriting.



(a)



(b)

Fig. 1. Effects of word contour information on human reading performance in four types of word patterns: machine font (sharp vs blurred) and cursive handwriting (sharp vs blurred). Contour condition were: 'Common', i.e. a frequently occurring pattern of ascenders and descenders; 'Unique', i.e., an uncommon word contour, and 'Flat', i.e., a word contour without ascenders and descenders. Reaction time is given in milliseconds (a). The blurring of words slows down the reading of words in general. Flat cursive words (i.e., without ascenders and descenders) are read significantly slower than words which are rendered in a machine font. The word recognition error rate (b) in percents of cursive script is unaffected by blurring if the word shape is 'Unique' in a word corpus, whereas reading of common and flat cursive words is deteriorated by blurring in comparison to clearly rendered script. Recognition rate of machine font is not affected by the blurring.

this degradation, whereas in cursive words with an infrequently occurring ascender/descender pattern the word recognition error rate is unaffected by the blurring: Note the error minimum for condition 'unique' in blurred cursive words. Blurring of machine font had no effect on accuracy, the curves are virtually overlapping in Figure 1b.

As a sequel to this study, we are currently exploring ways to determine the cursive script features used by humans in a more detailed way. In other studies, this has been done by manipulating the stimulus material, e.g., by masking parts of the text image with a pattern or by leaving out parts (rectangles) of the bitmap [6,7] (Figure 2).

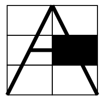


Fig. 2. Masking is done by making a rectangle of the bitmap black, such that a block-shaped part of the letter 'A' cannot be seen anymore.

Since it is not known in advance where the features are located, these masks will usually only partly overlap the 'actual' features used by the human reader. Another well-known method in reading research is the masking of words using random-pixel patterns. Mostly this affects whole words [8]. The disadvantage of such masking methods is that they severely influence the perceptual process.

In this study, a method is presented which allows the reader to indicate points of interest in the deciphering of handwriting patterns. Clicking with the mouse on a point in the image will trigger a local convolution of the image with a bell-shape kernel, enhancing the intensity of a handwritten trace, if present within the range of the convolution kernel. Consequently, there is no pre-imposed shape manipulation of the handwriting trace, other than its luminance on a CRT screen. The proposed process can be seen as a slow version of the normal perceptual process, in which visual information is processed during brief eye fixations, with fast saccadic eye movements between fixations.

In this paper, however, no strong claims are made concerning the actual neuro-perceptual process. The purpose of the method is to find out by which strategy an experienced human reading agent will solve the problem of feature extraction. The method is hypothesized to mimick the usually unobservable process of localized feature extraction in human reading by means of the externalized and observable 'mouse-clicking' behavior and can be viewed as a metaphor for a selective attention process. The advantage of this method over other potential approaches such as eye-movement recordings is its convenience. Although the recording of eye movement may seem a more natural approach, the goal of the experiment (detecting individual letter features) requires very large word patterns to be presented on screen, much larger than is the case in normal reading. In normal reading, the eye fixations cover syllables and whole words. Since the proposed method is slower than eye tracking, a time-pressure constraint is added to prevent elaborate cognitive reasoning from influencing the early perceptual process which is the actual focus of this study.

As a first test, the aim is to corroborate the findings from an earlier study [5] on the influence of the word contour on reading cursive. Predictions are that subjects will click frequently on the extremities in ascenders and descenders. The ultimate goal is to identify - for all letters in the alphabet - the zones of interest in human reading, thereby uncovering essential features in cursive script. It is also expected that the method reveals aspects of feature extraction as well as uncovering reading and segmentation strategies. For instance, many algorithms in handwriting recognition assume a strict left-to-right processing order, either in space or in time (or in both). From human reading studies, it is also known that final letters in a word may have an advantage to be recognized more easily [9]. It would be interesting to know to what extent the measured reading strategies will corroborate such assumptions and findings.

2 Method

Thirty-five subjects took part in the experiment. A data collection of 210 cursively handwritten words from different writers was divided in five data sets of 42 words. This set size was used to reduce fatigue effects. Such a data set was presented to seven different subjects, i.e., there were five groups of seven readers each. The subjects were seated in a dimly lit room under constant lighting conditions. They were presented with a grey CRT screen on which handwritten word images appeared with an intensity (luminance) just above the background level. The distance of the eyes to the monitor was constrained to 50 cm by a horizontal string in front of the face, stretched below eye-height at the border of the table on which the monitor was placed, to prevent subjects from leaning forward to the screen.

Care was taken such that words were rendered with a low luminance at the beginning of a trial such that they were illegible. By clicking on a location on the screen with the mouse pointer, subjects could light up an area with a luminance curve which radially tapers off towards the grey background level. The effect is a bell-shaped function of intensity which only affects the handwritten trace, not the background. The luminance effect of a click has a view angle of about 0.6 degrees, the maximum horizontal view angle of words presented on the screen is 18 degrees.

Figure 3 shows a word which appears in a central window of the screen. The word which is written is *popcorn* and several clicks have already been made.

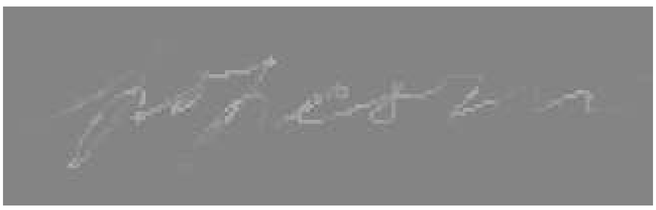


Fig. 3. A word in the central window of the experimental environment: popcorn. Several clicks have been made, to make the word more visible.

Subjects were asked to write down the correct word as soon as possible, with a minimum of clicks during the reading process. Subjects operated under time pressure to prevent too much conscious reasoning from taking place. In this way it is expected that it is mainly the perceptual process which plays a role in the word classification. To further enhance the difficulty of the task, words with a low frequency of occurrence in normal written language are used. The word list contains all letters of the alphabet and has been designed to cover as much digrams as possible. The word list is bilingual (English, Dutch). It was on-line recorded with an electronic paper device (Wacom PL100V), with a sampling rate of 100 Hz (resolution 1/50 mm, accuracy 0.1 mm). Connected-cursive handwriting by ten different writers was used to ensure the presence of a wide range of cursive styles.

3 Analysis

The time t (in ms) of the moments of clicking during a trial, and the XY coordinates of the points of clicking (in mm) are stored. The on-line recorded words have been annotated at the character level, such that for each click, the nearest letter in the stimulus word could be determined. The following dependent variables were studied: vertical position of clicks in a letter with respect to the handwriting baseline, normalized number of clicks per letter, and the number of clicks per letter position in words.

Each letter in the lexicon was labeled by hand and divided into different allograph-groups. For each group we generated an allograph prototype, on which the click-density pattern could be projected. In order to achieve this, the allograph replications were normalized to a fixed origin and a root mean square (r.m.s.) radius of one, applying the same transformation to the XY values of the corresponding clicks. This latter procedure allows for the detection of singularities or points of interest to a human reader.

4 Results

After the experiment, the subjects were asked to read the words in normal image conditions with clearly visible script. In the dimmed condition, the recognition rate after enhancing features by clicking is 62.4%, whereas in normal visibility, the human word recognition rate is 87.9%, similar to what has been found elsewhere [10] recently. Note that the latter performance is obtained after exposure to the words of the lexicon. These findings indicate that this was not a trivial task for the human readers.

Figure 4 shows the vertical position Y of the clicks on lower-case letters, taking the handwriting baseline as zero. The values are sorted in order of increasing Y, yielding the given distribution of letters on the horizontal axis. As can be seen, the letters with descenders are concentrated on the left, the small, corpus-sized letters are in the middle, whereas the letters containing an ascender are on the right of the distribution. The <*f*>, which in most cursive writers has both a descender and an ascender stroke in cursive handwriting is located between corpus-sized letters (most hits were in the middle zone of this letter).

Figure 5 shows the distribution of average number of clicks on a letter for the lower-case letters of the alphabet. Note that the number of clicks on each letter has

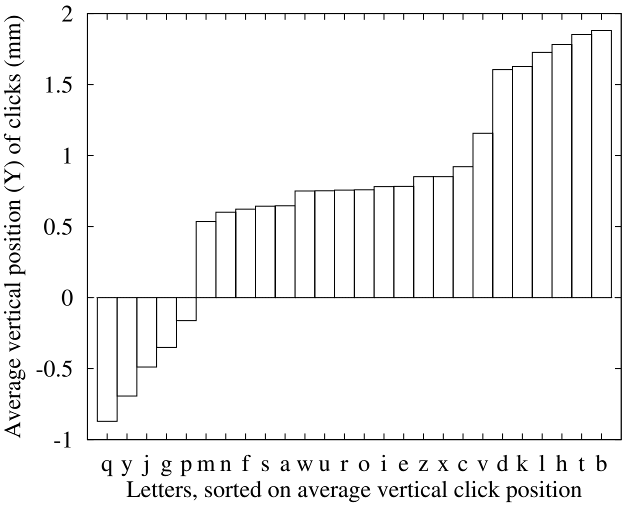


Fig. 4. Average vertical click position. Y_{click} in [mm] per letter, sorted in ascending order. There are three zones: descenders (left), corpus-sized letters (middle), and ascender letters, which appear to cluster on the right side. The baseline is located at zero mm. Note the negative click positions for descenders, and the discrete jump between the Y_{click} values of v and d at a height of about 1 mm.

been normalized by dividing the number of clicks on a letter i by the number of times this letter i appears in the word list. This normalization is necessary to allow for a comparison of the click density between letters.

The list of letters shown on the x-axis in figure 5 is sorted in increasing number of elicited clicks. The average normalized number of clicks on corpus-sized letters (*aceimnorsuvwxz*) is 10.6, whereas the average number of clicks on ascender/descender letters (*bdfghjklpqty*) is 12.8, which is statistically significant (t test, $p < 0.05$). As can be seen, there is a concentration of letters representing vowels on the left.

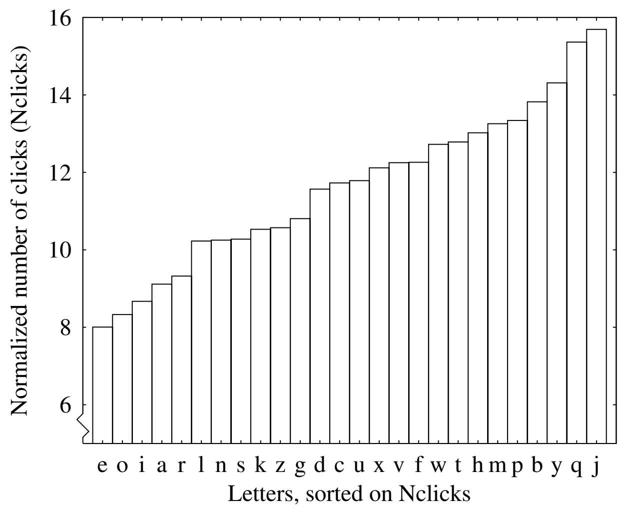


Fig. 5. Normalized number of clicks on each letter. Note the concentration of vowels on the left.

Figure 6 shows the distribution of clicks over the word pattern, from letter $-N/2$ to $N/2$, where different curves are shown for words of four up to 10 characters. Most 'clicks' are on the first and last letters of the word. There is no significant difference in the distribution of clicks between the first and second half of the word.

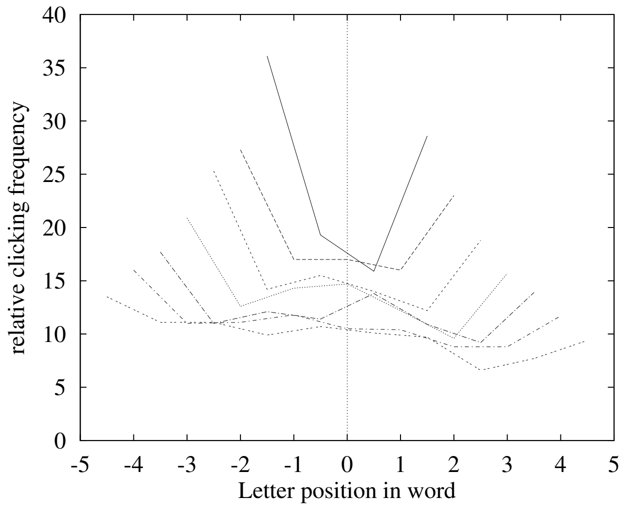


Fig. 6. Relative clicking frequency per letter position in words (0 represents the word center), for words of four to 10 letters.

The number of elicited clicks is not a simple function of the amount of letter pixels, letter height or letter width: Even when combined as predictors in a linear model, these variables explain only 40% of the variance in the number of clicks. Furthermore, it was checked whether the first and last letter elicited a higher 'clicking rate' due to a strategy effect: It is possible that the human readers only click on the first and last letter in order to locate the beginning and end of a word in the early stage of this 'reading' process, followed by a few clicks in the middle of a word. Inspecting the distribution of clicks on the word over time, however, we find clear indications that the subjects enhance the luminance of word section by clicking on the words, proceeding from the spatial left to the right in time (Figure 7), apparently increasing the number of clicks at the tail of a word.

Detailed analyses of the click density pattern reveals the actual distribution of clicks within allographs. In figure 8 are five examples of the kind of results we obtained. The allographs *jI*, *lI* and *yI* all three have comparable loop shapes, yet we see in the figure that the subjects have indicated different points of interest. For *jI* and *lI*, the crossing seems most important, but for *yI* the sharp edge is more important (more clicked on) than the crossing. Allographs *aI* and *aIII* are both examples where clear features emerge from the clicking densities: Allograph *aI* shows the importance of vertical strokes, *aIII* the importance of high-curvature points (note that the allographs are cut out of their word context in these figures). To get a good impression of the densities, it may be required to look at this figure from some distance.

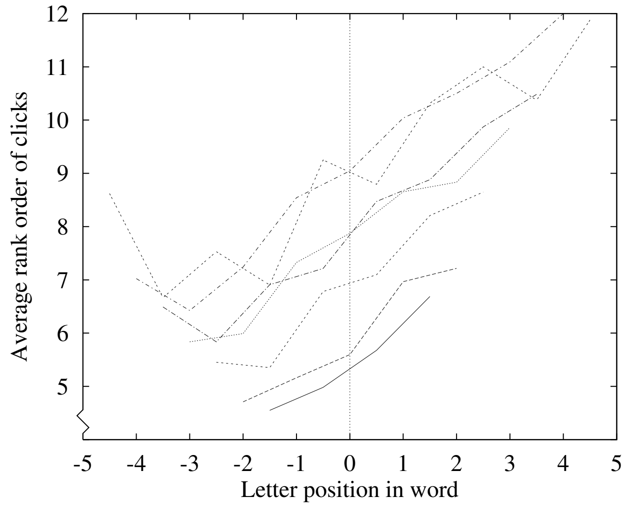
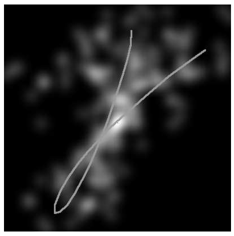
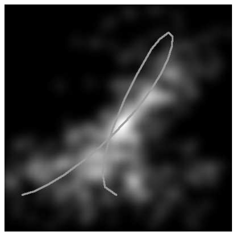


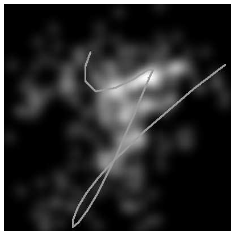
Fig. 7. Average sequence number (rank order in time) of the feature-exposing clicks during the word-reading trial period, as a function of letter position in the word. The central letter is at position 0. In general, the 'reading process' proceeds left-to-right in both space and time. A small exception are the longer words, of which the first letter may be occasionally missed, followed by an excursion to the left (cf. top-dotted line, at point (-4,8)).



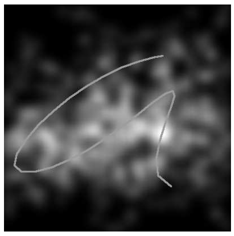
(a) jI (13,194)



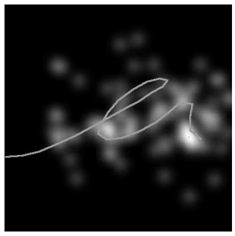
(b) lI (57,583)



(c) yI (23,415)



(d) aI (112,1276)



(e) aIII (14,121)

Fig. 8. Click distribution in allographs. The roman number identifies the allograph. Frequency of occurrence (N_o) of the allograph in the dataset and number of clicks (N_c) are given as (N_o, N_c). <*jI*> and <*lI*>: these patterns show that the crossing in a loop may attract the attention. In the pattern obtained for <*yI*>, the clicking density is highest at the point of high curvature at the top of the descending stroke. <*aI*> is an example of high clicking density elicity by a downstroke in this allograph. Finally, <*aIII*> reveals that points of high curvature may attract the attention of the human reader.

These examples of clear feature locations emerging from the clicking density are assumed to play a role in the selective-attention process and in the subsequent classification of the letters in the human reader.

5 Discussion

We have found that shape information (presence of ascenders and descenders) is important for recognition. This is a confirmation of earlier work [5], which obtained the same results with a completely different experiment. Although this would seem natural to many researchers in the area of automatic handwriting recognition, it is very difficult to find similar clear signs of word contour usage in the human reading of machine-font characters. Furthermore, we found that characters representing the vowels seem less important for recognition than consonants. This can be due to the fact that vowel characters do not contain many features which are essential for the word recognition process, with its strong dependence on the mental word lexicon: A word can often be read if the vowels are left out (*hndwrtng*). Some languages, e.g. Arabic, virtually omit the vowels in the script. Furthermore, it was found that both the first and last letter of the word are very important for the recognition process.

The high clicking rate at the rightmost part of words is an unexpected finding. Temporal data show that it is not a fencing or detection strategy in the early 'reading' stage. It seems more likely that the subjects actually need the last letter(s) for the word recognition process. This observation is interesting, because during left-right (LR) word search in character-hypothesis space by many automatic recognizers it can be observed that the number of word hypotheses keeps growing steeply, only to taper of to a smaller list of word hypotheses when the final, rightmost letters are taken into account. It is likely that in automatic handwriting recognition based on LR word search, a sub-optimal dehooking process of the word tail and an improper handling of trailing t-bar crossings and <*i*> dots will have a disproportionate deteriorating effect on word recognition performance.

The results of the click-density patterns show that our method of determining features by observing a paced-down human reading process turned out surprisingly well. On the one hand, the results confirm known phenomena, such as the importance of vertical strokes [11] (for example allograph *aI*) and crossings (for example allographs *jI* and *lI*). On the other hand, we also have found new features which subjects seem to use for recognition. The most important part of the writing curve seems to be where the writing speed has reached a minimum and curvature reaches a maximum[12]. Also, curled endings of a final stroke often attract the attention of the reader's eye. Some handwriting recognition approaches make explicit use of the high-curvature points in handwriting for segmenting the handwriting trace into velocity- or curvature-based strokes [13].

6 Conclusion

We have introduced a new method for the determination of letter features in cursive handwriting. The method not only confirms already known phenomena such as the importance of crossings and down-strokes, but also reveals new information about important features. Although the procedure is not guaranteed to represent the natural reading process, it does convey information about the strategies and features that an experienced human reader uses to solve the task. Apart from using the proposed method for determining important features in handwriting it may also be used for determining important features in machine fonts and in images and pictures [14]. It is expected that the behavior of automatic handwriting recognition algorithms can be improved considerably by using features which are similar to those used by the human reader, especially if the algorithm itself is also based on knowledge of human reading [15]. It is to be expected that the unavoidable errors made by such script recognition systems will be ultimately less erratic and counterintuitive to the users than is currently the case in the behavior of handwriting recognition algorithms.

References

1. Paap KR and Newsome SL (1984) Word shape's in poor shape for the race to the lexicon. Journal of Experimental Psychology: Human Perception and Performance, 10(3):413–428
2. Mayall K, Humphreys GW, Olson, A (1997) Disruption to word or letter processing? The origin of case-mixing effects. J. of Exp. Psych.: Learning, Memory, and Cognition, 23(5):1275–1286
3. Corcoran DWJ, Rouse RO (1970) An aspect of perceptual organization involved in reading typed and handwritten words. Quarterly J. of Exp. Psych., 22:526–530
4. Jacobs M, Grainger J (1994) Models of visual word recognition - sampling the state of the art. Journal of Experimental Psychology: Human Perception and Performance, 20(6):1311–1334
5. Brassé B (1991) De invloed van de globale woordcontour tijdens visuele woordherkenning: een vergelijking tussen drukschrift en handschrift. (The influence of global word contour during visual recognition: a comparison between machine print and script.). Technical Report Report 91-NICI-04, Nijmegen University, The Netherlands, 1991.
6. Suen CY (1983) Handwriting generation, perception and recognition. Acta Psychologica, 54:295–312
7. Suen CY, Guo J, Li ZC (1992) Computer and human recognition of handprinted characters by parts. In From Pixels to Features III: Frontiers in Handwriting Recognition, North-Holland, pp 223–236
8. Snodgrass JG, Poster M (1992) Visual-word recognition thresholds for screen-fragmented names of the snodgrass and vanderwart pictures. Behavior Research Methods, Instruments & Computers, 24(1):1–15
9. Humphreys G, Evett LJ, Quinlan PT (1990) Orthographic processing in visual word identification. Cognitive Psychology, 22:517–560
10. Barrière C, Plamondon R (1998) Human identification of letters in mixed-script handwriting: An upper bound on recognition rates. IEEE Trans. on System, Man and Cybernetics - (B), 28(1):78–82
11. Powalka RK, Sherkat N, Whitrow RJ (1995) Recognizer characterisation for combining handwriting recognition results at word level. In Proceedings of the 3rd. Int. Conf. on Document Analysis and Recognition (ICDAR '95), pp 68–73
12. Attneave F (1954) Some informational aspects of visual perception. Psychological Review, 61(3):183–193
13. Schomaker LRB, Teulings HL (1990) A Handwriting Recognition System based on the Properties and Architectures of the Human Motor System. In Proc. of the Int. Workshop on Frontiers in Handwriting Recognition (IWFHR), Montreal, CENPARMI, pp 195–211
14. Vokey JR, Baker JG, Hayman G, Jacoby LL (1986) Perceptual identification of visually degraded stimuli. Behavior Research Methods, Instruments & Computers, 18(1):1–9
15. Coté M, Lecolinet E, Cheriet M, Suen CY (1998) Automatic reading of cursive scripts using a reading model and perceptual concepts. International Journal on Document Analysis and Recognition, 1:3–17