Handling within-writer variability and between-writer variation in the recognition of on-line handwriting. *

Lambert Schomaker
NICI, Nijmegen Institute for Cognition and Information
University of Nijmegen, P.O.Box 9104
6500 HE Nijmegen, The Netherlands
E-mail: schomaker@computer.org

1 Introduction

Although the performance of current automatic recognition algorithms of on-line handwriting has much improved in recent years, there are still many problems with the actual application of these systems. It appears that the step from academical experiments to real-life use of such algorithms, in, e.g., portable pen computers, is still difficult. What is particularly intriguing is the fact that reported academic (and commercial) recognition rates usually are 10-20% overestimated. When such systems are given the real test, i.e., use by *any* writer, in a realistic application such as note taking during lectures, their performance drops sharply. One reason lies in the fact that for on-line handwriting, only limited training databases exist. A project is currently running to alleviate this problem: UNIPEN (Guyon & Schomaker, 1994). Although the availability of huge databases for system training and development potentially improves the performance of existing algorithms due to the wider coverage of handwriting shapes, it is very likely that many algorithms are not well fit to handle the case of an infinitely large training set. Neural network-based approaches, but also approaches based on hidden-Markov models both run the risk of satiation, where the system yields an average but incomplete representation of all possible handwriting shapes. Similarly, brute force matching methods run the risk of becoming computationally impractical, when all possible character shapes (allographs) have to be considered. This study is directed at the development of procedures to obtain an insight in the underlying variation of shapes within large quantities of handwriting data from several writers. At this stage, it is useful to make a distinction between two source of variation in handwriting shapes:

1. Between-writer Style Variation
2. Within-writer Variability

First, there is the large variation in handwriting styles over individuals. Even within the main groups of handwriting styles (i.e, isolated hand print, connected cursive, and mixed cursive), a spectacular variation in allographs (form variants) for a given letter can be observed (Figure 1). This is mainly dependent on the writing method taught at primary school, but also on personal preference, the copying of style variants from peers in adolescence, and later in age, by the amount of writing experience. The second source of writing variation, or rather, variability, holds for a given writer, and expresses itself in subtle, low-frequency movement noise, slant and size variability, as well as in a varying choice of allographs. Some writers will even jump from hand print to a mixed-cursive and a fully connected cursive style at a single writing occasion. For

*NICI Technical Report, presented at the 7th IGS, London Ontario, 1995. Supported by Hewlett-Packard, Bristol.

the automatic recognition of cursive handwriting, the problem of variation and variability has generally two effects. First, a large number of allograph shapes must be represented in a recognition system, sometimes leading to excessive computation time in character and especially in word matching. Second, the presence of writing styles of other writers in the recognition system will yield unlikely character matches, inappropriate for the writer currently using the system. These observations are relevant to a broad range of recognition methods. It should be noted, that in the recognition of machine-font characters, the knowledge on character-shape families is very exact. Surprisingly (to the researcher in handwriting recognition) the shapes have different names even for font families with only a few differing pixels for a given letter in the alphabet. This information is explicitly used by current OCR systems (Figure 1). It would be conducive is knowledge on allographs in handwriting would be made explicit in a similar fashion.

a) **AvantGarde Book, 24pt**
these letters look very much like
b) **these letters look.**
Helvetica–Bold, 24pt

Figure 1. In machine-font recognition (OCR), explicit knowledge on font families is known, even for fonts differing only a few pixels per character. The phrase **"these letters look"** in AvantGarde Book, 24pt (a) as contrasted with the same words in Helvetica-Bold, 24pt. (b)

In order to be able to ultimately alleviate the problem of variation and variability, two experiments have been performed to measure these phenomena.

In experiment A, targeting the between-writer variation, the idea is pursued that writing style can be defined as "the set of typical strokes" used by a given writer. In this view,the identification of allographic style variation is postponed, and a bottom-up analysis of typical stroke usage histograms is done in order to identify known writers. This approach has the advantage that it can be performed automatically, without supervised training such as the manual labeling of allographs. From a pilot study, it appeared that writers can be very well recognized with this approach, on the basis of twenty on-line recorded handwritten words. In the current study the idea is addressed more thoroughly and will be applied to a larger writer group. The results may be extended to identification of generic sub-styles.

In experiment B, which is on within-writer variability, measures of variability will be computed for a number of global, allograph and stroke parameters. Part of the data will be composed of handwriting samples, collected at several different occasions spaced two weeks apart, for a number of writers. Another part of the data will be composed of handwriting collected under a number of different conditions (words vs sentences, and dictation copying vs copying by reading). Results will be interpreted in terms of applicability for use in the automatic recognition of on-line handwriting, and in terms of potential consequences for handwriting production models. Preliminary results have shown that dictation copying leads to more variable handwriting, which can possibly be attributed to cognitive phonemic-graphemic conversion overhead.

The necessity for cataloging Western handwriting styles becomes more and more apparent as on-line handwriting recognition algorithms currently reach an asymptote in their performance, and a limited generalization from laboratory training set to real life conditions is observed. Although the algorithms as such still need to be refined, and an optimal approach has not as yet been identified, performance improvement is most likely to result from the availability of much larger training sets of on-line handwriting data than is current practice.

Indeed, in the comparable fields of speech recognition and optical recognition of handwriting the situation is different. The speech recognition area already has a large, commonly accepted test bed for evaluating recognizers, like the TIMIT database. In the optical recognition of handwriting, the main international post companies all have a huge base of scanned texts from actual mail envelopes, and the continuous flow of data is regularly sampled to retrain recognizers in order to capture trends in change of styles. Consequently, the research area of off-line optical recognition but especially that of speech recognition is in a more advanced technological state than is the case in on-line handwriting recognition.

In the HP/NICI collaboration project, the problem of handwriting style has been analyzed as to consist of two components:

1. Between-writer Style Variation
2. Within-writer Variability

Ad 1. In Western culture, a huge variation in writing styles exists. Between different European countries there are clear style differences. Even within a country, there are style variations (Figure 2) caused, e.g., by differences in writing methods at primary school. As a consequence, there may also be clear differences between writers from different school generations.

Apart from work in forensic handwriting analysis (e.g., the German B.K.A. system FISH), there exists no catalogue of Western handwriting styles and little is know about algorithms to calculate quantitative measures which can be utilized in on-line recognition systems.

Ad 2. Apart from differences between writers, however, there is also the phenomenon of variability of handwriting within an individual writer. Four types of variability exist:

(a) geometric variability without change in the "topological" characteristics of characters; (b) omission of strokes (fusion) due to fast or careless writing; (c) insertion of strokes or ligatures, in elaborate writing or in the case of hesitations or spurious pen movement; (d) letter shape (allograph) variability due to stylistic choice.

The first type of variability (a) comes from the neural noise in the human motor output system, and leads to geometric variability in the form of slant and roundness deviations per stroke, essentially however, preserving the "topology" of the characters (Figure 3).

The second type of variability (b), stroke fusion, can theoretically be explained as follows. Let us assume that we can make a distinction between a central pattern generator and a pipeline of transforming filters, initially being neural, but the final filter being composed of the biomechanical effector system. The filtering properties of the output channel as a whole are essentially of a low-pass nature. The observed bandwidth of handwriting is about 10 Hz (Teulings & Maarse, 1984). According to the minimized-jerk theory (Flash & Hogan, 1985), the movement trajectory is generated on the basis of the constraint that so-called "via points" are reached (in our case, topologically important points in a single character), and that the rms value of the first derivative of acceleration is minimized. The pattern generator plans the sequence of x,y via points. Under conditions of reduced mental concentration or speed requirements, the central pattern generator (partially) omits some via points in its output, leading to fused strokes, yielding less prominent character details (Figure 4).

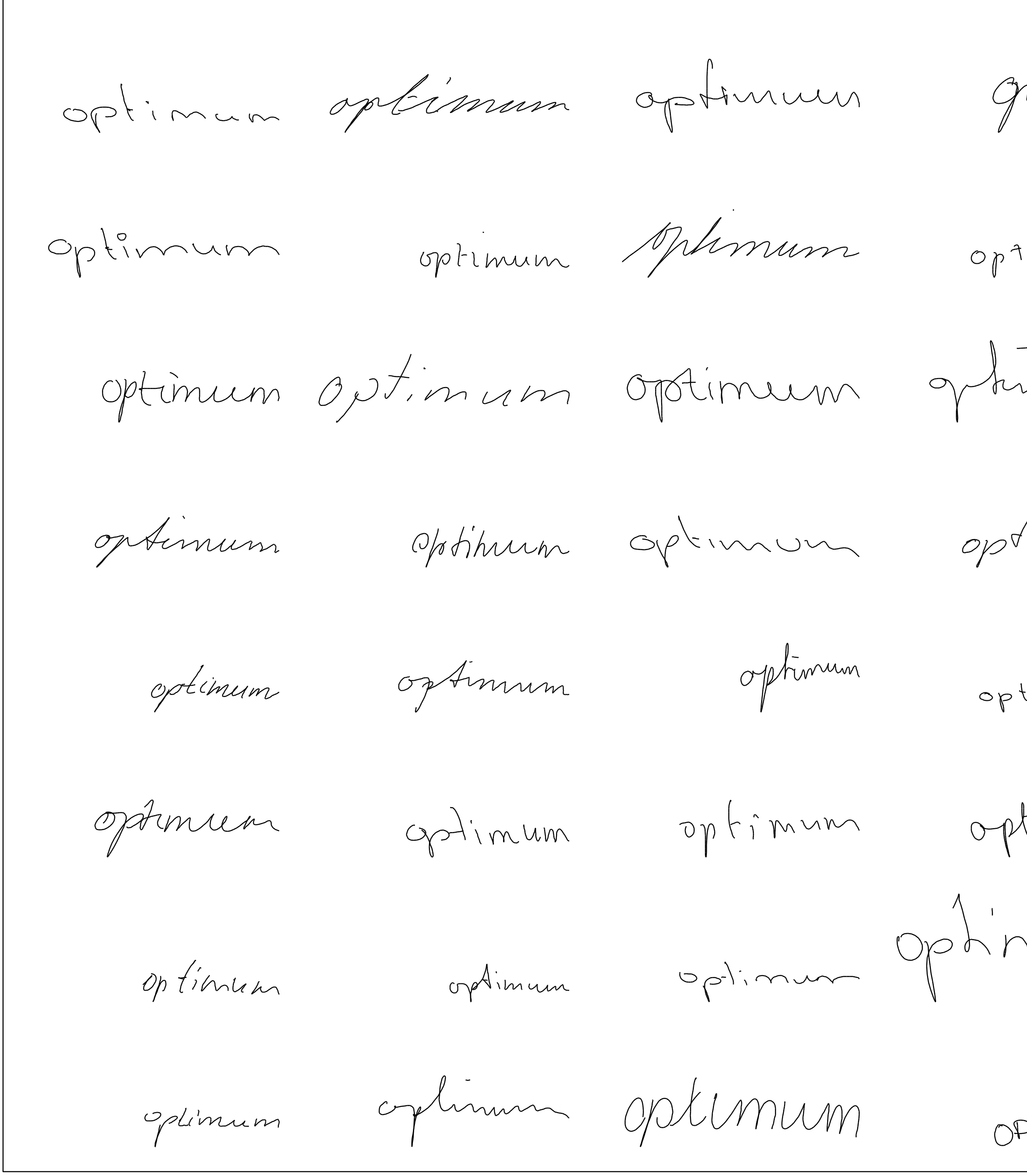

Figure 2.
Style variation between writers. Different samples of the word <*optimum*> for 32 different writers.

Figure 2.
Within-writer variation: the case of limited human-motor noise. Several samples of the word <*algebra*>. Rows represent eight different writers, the four columns represent different replications of the word, written at different points in time. Words written in column 1 vs 2 (and 3 vs 4) are separated maximally 2 hours in time. The two leftmost columns (1 and 2) are separated minimally two weeks in time from the two rightmost columns. In row 1, (cursive) the loop in the <*g*> is missing, whereas the other three replications of <*g*> are looped. In row 2, (mixed cursive) the pen is lifted at different points in different replications. A closed and three extremely open variants of <*a*> are produced. In row 8, (mixed cursive) two allographs of the <*r*> are used.

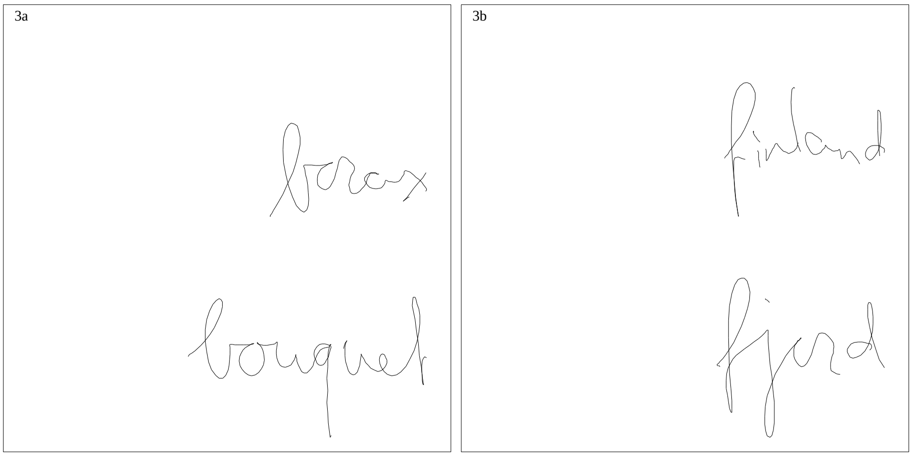

Figure 4. Fusion of strokes, differences within a single writer. (3a) The words <*borax*> and <*bouquet*> show that the <*or*> transition leads to a fusion of the last stroke of the <*o*> into the connection stroke with the <*r*> in <*borax*>, whereas the <*o*> in <*bouquet*> is neat and complete. A similar phenomenon occurs in the <*ax*> transition in <*borax*>.
(3b) The word fjord shows a similar stroke fusion in <*or*>.

The third (c) form of within-writer variability is caused either by similar high-level processes as in (b), this time however inserting strokes at will, or, alternatively by interruption of the central patterning process. The latter can be self-induced, when the writer thinks about the formulation of the text to come. This phenomenon is called "phonemic-graphemic interference". The phonemes of words-to-come are activated subliminally (i.e., without giving rise to speech musculature activation), but with sufficient levels of activation to produce a premature spelling process activation. The resulting allograph "breaks into the current motor output buffer". Other causes of inserted erroneous strokes are external events, such as loud noises, doors opening, phones ringing etc., after which the writing process resumes.

The fourth (d) form of within-writer variability originates at a higher, cognitive level in the human writing system and has to do with the choice of letter shapes (allographs). For example, it is often observed in user-trainable systems that writers enter different shapes in the training stage compared to the letter and word shapes entered in the actual use of an application. Within a single writer, there may even be a seemingly random choice of styles as different as isolated hand print and connected cursive.

Both components of variability in handwriting: Between-Writer Style Variation and Within-Writer Shape Variability can only be handled effectively by on-line recognition algorithms if more is known about their statistics: Which variables are essential, and what are their distributions, and can we identify clusters of generic writing styles?

In order to approach this problem in the areas of on-line recognition of handwriting, the HP/NICI collaboration-project team has designed a data collection setup fulfilling a number of purposes, as described in the next section.

2 Criteria for an on-line handwriting data set suitable for addressing the variability problem

The data set to be collected:

1. must capture style variation among writers,
2. must capture style variability within a writer, as measured at occasions sufficiently spaced apart in time,
3. must be large enough to allow for a number of large-scale training/testing experiments,
4. must be compatible with the UNIPEN project, so that data from other institutions may also be used in such massive training and testing,
5. must be of high quality as regards the signal properties, since deteriorated signal conditions can easily be imposed post hoc.

2.1 Additional constraints: input unit scope

The data collection is WORD-oriented, since recognizers at both HP and NICI are based on isolated word recognition. Also, this is the input chunk size currently handled by most free style or connected cursive recognition systems. The **letter** level is only suited for isolated hand print and digit data. The **sentence** level and higher (**paragraphs**, **pages**) impose additional word segmentation problems which are difficult to handle at the moment. It is not completely possible to compute word segmentation on the basis of bottom-up features like white space or ink clustering: Often lexical or even syntactical top-down information would be necessary to disambiguate here. In many applications, however, the word-based input is already useful, especially if recognition speed can be fast enough to not disturb the human word production process ("train of thought") (Nakagawa et al., 1993). The WORDS will consist of lower case characters.

2.2 Additional constraints: word lexicon

The elements of a word list in handwriting collection setups is usually a subject of hot debate due to the large number of possible criteria for inclusion (size, word length, character content, digram content, trigram content, linguistic frequency of usage, etc.). In the collection setup, two basic constraints were chosen, sacrificing some other criteria:

2.2.1 Bilinguality

The list must be bilingual in the sense that the same list can be written by Dutch and English writers. This allows for the incremental collection of words in both Nijmegen and Bristol. It will ensure that the Dutch writers will not feel uneasy writing a foreign language.

2.2.2 Maximized digram coverage

In connected-cursive and mixed-cursive handwriting, the current character shape is determined by both predecessor and successor. The connecting strokes come from a previous character, retaining effects from the starting position and the angular velocity (clockwise, sharp, counter-clockwise), and may exert an effect on the first strokes of the current character itself. Similarly, the anticipation of the next character may lead to distortions of the final stroke(s) of the current character. To obtain a reliable overview on character production strategies, as much digrams from the 26x26 transition matrix must be present in the word list.

Actually, there are 27 symbols, including the space symbol (identifying Begin-Of-Word and End-Of-Word conditions).

In order to build a word list that fulfills the aforementioned criteria, the following approach was taken.

2.2.3 Steps in determining the word list

- Word List 1: 50k Dutch words.
- Word List 2: 50k English words.
- These two word lists were ran through Unix comm, yielding a list with 3251 words common to both languages.
- As the resulting list was too large for the data collection process, it condensed with a dedicated program in C which created a subset of words with the criterion of maximum digram coverage. This means that all (27x27) digrams present in the input list will be present in the output list. The program is based on stochastic optimization, iteratively picking a word from the input list with a low probability, and only adding it to the output list if it contains new unseen digrams. This was done several times, choosing a final list which was acceptable (decency, not too difficult to spell, etc.). The resulting word list contained 210 words. Due to the selection algorithm, the words are slightly longer than average English words.
- A number of words was manually added because of their interesting (but low frequent) digrams. An example is the <*x-y*> digram in "xylophone". For this word, the English spelling was used which is more acceptable to Dutch writers than "xylofoon" would be for English writers. The final list consists of 210 words (Appendix I).

The word list contains many international concepts (e.g., "algebra"), geographical names, technical terms, latin-origin words, french-origin words, as well as words which happen to be spelled the same in both languages, but may have a different meaning ("trekking"). After the writing sessions, the subjects were asked from which (unmentioned) language they thought the word list was, and also they were asked to mark words which they thought were difficult to write. The list appears to be of medium difficulty, and there were no specific complaints by the subjects.

3 Recording Setup

Since a representative "real-life" application does not yet exist, it was decided to collect words in a visually prompted word setup with a provision for rewriting words the subject considers badly legible him/herself. In such a case the subject would tap a *Cancel* button on the screen, instead of the normal *Next* button. Words are randomized on each session. Writers sat at a table in a room with dimly lit fluorescent lamps to prevent glare from the Wacom PL-100V LCD screen. The Wacom was placed on a normal desktop in an orientation preferred by the subject. A separation panel was placed between experimenter and subject to prevent additional stress or performance pressure which often develops in experimental setups. Subjects are eager to please experimenters, and sometimes weary of hidden motives (intelligence or personality tests). For our purpose it was important that writers used **their own**, i.e., their mostly-used handwriting, rather than a style they thought was acceptable. There was an introductory text on a sheet of paper, and writers were allowed to get accustomed to the setup by writing 20 habituation words.

4 Session Schedule

The subject came to the lab three times (Sessions), spaced two weeks apart (Table 1). At each Session, two Sets of the 210 words were produced, yielding six Sets (totalling 1260 words written per writer). Within a Set, the writer was allowed to pause after 100 words.

Session 1		*Session 2*		*Session 3*
Set 1 Set 2	→ (2 weeks) →	Set 3 Set 4	→ (2 weeks) →	Set 5 Set 6

Table 1: The schedule of recording sessions. A 'Set' refers to the writing of a single list of 210 words

Data from 35 subjects has been collected, writers producing the word list 6 times each.Subjects were asked if they were available for later collection occasions.

5 Recording Software

The recording software consists of a Visual Basic application (PLUCOLL) and a DLL package written in C (PLUTO) for the actual sampling of the pen-tip coordinates. The output consists of individual UNIPEN-format files per word. (the .INK files), as well as a writer description and a setup description file, written to the local hard disk on the PC. After each session the collected .INK files and information files are combined in a single UNIPEN file for a set (e.g. SET1.DAT). This is done by the program UNIWRAP, which produces a UNIPEN file on the basis of a checklist of constituent file names. PC-NFS was used for Unix disk access (the UNIWRAP output files are written to a remote disk on a HP 9000/735 workstation.

Operating system: DOS 6.2, Windows 3.1, Windows for Pen Computing 1.01a,

Application: data collector written in Visual Basic V3.0

Network: PC-NFS V5.0a

Tablet Driver: MS Windows for Pen Computing V1.0 PENWIN.DLL,
plus a custom Borland C++ V3.1 DLL routine calling GetPenHwData

6 Recording Hardware

PC: IBM 486SLC2-66 MHz motherboard, 4 MB.

Network: 3COM 3C509 Ethernet adaptor.

Tablet: Wacom PL-100V in landscape orientation

Writing box: width=117 [mm], height 25 [mm], with a dotted line at 11 [mm] from the bottom. A progress bar was presented at the top. Words to be written are presented in a box 20mm above the writing box, left justified, Prompted word font: MS Serif, 18pt.

Inking: Black ink on white background, ink width of 1 pixel.

Sensor: Electromagnetic, wireless pen (transponder)

Pen: Untethered Pen, Tip Switch

Sampling mode: Continuous, equidistant in time, during PenUp & PenDown

Sample rate: 100 [Hz]

Resolution: 0.02 [mm/unit]

Coding Category	Explanation
spelling	This is the worst possible category: human readers read a different word from what has been written.
stroking	This category refers to fused or omitted strokes
punctuation	Refers to unsollicited punctuation/diacritics
capitals	lower case characters were sollicited only
disconnected	as in <*cl*> or <*ol*> denoting <*d*>, with a very clear white space in between two components.

Table 2: Annotation categories of special, non-optimal word quality cases

Accuracy: 0.1 [mm]

Width: 192 [mm]

Height: 143 [mm]

7 Subject Group

In this data collection setup, we tried to avoid the usual population of co-researchers and students. The target group was older than 25 years, and a number of professions in which writing is a usual activity was included. This was done by recruiting people through a newspaper advertisement in a medium-sized Dutch paper. The average age is about 30 years. Handedness L/R is distributed proportional to the whole population (approx 1 in 10 left handed). The average computer experience is 5.5 years, this is partly due to three subjects having more than 10 years experience. Two subjects have no computer experience. About half of the subjects have university training, the other half having various backgrounds. The profession was mainly from "Services" (other categories were: Medical, Industrial, Education, Office, Technical, Research, None). The majority of the subject wrote mixed cursive, according to their own judgment. The others claimed to write cursive (They were shown four words samples from the categories Block print, Handprint, Mixed cursive, and Cursive).

8 Data Annotation

The UNIPEN program UPVIEW was used to annotate the SETx.DAT files word by word. By clicking on a word box in UPVIEW, a flat text editor appears with on the first line the label of the word that should have been written. The annotator can place remarks in this file. The following categories of special, non-optimal word quality cases were defined:

The annotation appears in individual files, e.g., the fifth word of set1.dat will be annotated in a separate file set1.dat-segment-4.log More details are given in Appendix II.

9 State of the Work in Progress

Currently, individual character labeling is performed interactively. Words are sent to the NICI script recognizer. The recognizer is set to a strict recognition mode, i.e., individual characters must have a posteriori probability of $p > 0.05$. Furthermore, all individual characters in a word must be identified, yielding a contiguous letter path representing the correct word, never missing more than two strokes between two letters. If the word is recognized, the resulting labels are stored (in word*nnn*.lbR files, where "R" stands for Recognized). If a word is not recognized, the operator labels all the characters in a word manually, including the connecting strokes. If characters are illegible by human or if the words are misspelled, the corresponding characters are not labeled. The labels produced by the human operator are stored in separate files (named word*nnn*.lbl). In order to maintain a consistent labeling strategy, there is regular supervision on the process.

10 References

Flash, T., & Hogan, N. (1985). The coordination of arm movement: An experimentally confirmed mathematical model. Journal of Neuroscience, 5, 1688-1703.

Nakagawa, M., Machii, K., & Kato, N. (1993). Lazy Recognition as a Principle of Pen Interfaces. Conference handout (nakagawa@tuatg.tuat.ac.jp).

Teulings, H.L. & Maarse, F.J. (1984). Digital recording and processing of handwriting movements. Human Movement Science, 3, 193-217.

11 Appendices

In Appendix A, the list of used words is shown, dubbed the NLUK-210 list. Also the digram frequency table is given for this word list.

In Appendix B, some basic statistics of a subset of the collected data are shown, such as slant, and number of pen-down pieces. Look at the GrandMean, which is the average of the writer averages over each 210-word set.

Appendix C summarizes the database quantities and the state of the data.

In Appendix D, ficticious writer names are shown which will be used to identify these sets in the future. In the development of knowledge on style clusters, it will be easier to refer to such styles using these names (as a kind of "font" name).

Appendix F shows the correspondence between what writers thought was their handwriting style, and a simple measure of "connected-cursiveness", i.e., the average number of pen-down ink pieces per word ($Npiece$), for each writer. Indeed, writers who claim to write cursive, have the lowest average values of $Npiece \approx 1.8$, whereas writers claiming to write handprint yield an average of $Npiece \approx 8.6$.

A The 210-word NLUK list

abdomen
abstinent
adherent
adjunct
advocate
afghanistan
album
aldehyde
algebra
alluvium
alp
amanuensis
analyst
anecdote
angst
antecedent
aorta
appendix
aqua
arcsin
auschwitz
backup
badminton
bangkok
batik
bauhaus
bazaar
bhagwan
bijouterie
bladder
bobby
bodyguard
bolster
borax
bouquet
boutique
bradford
breakdown
brisbane
budget
buffet
byte
calcium
charisma
checklist
chevron
chloride
cockpit
cocktail
colonnade
comfort
concubine
conjunct
copywriter
cornwall
corps
cowboy
crawl
croquet
cycle
czerny
darwin
dashboard
deadline
debugger
dejeuner
delhi
delinquent
deodorant
diagnose
disjunct
dixieland
dizzy
dozen
drink
edelweiss
entertainment
equilibrium
equipment
essay
excellent
exodus
export
extract
exuberant
fascist
feedback
finland
fjord
flipflop
frankfurt
fuchsia
genre
gladiator
god
guyana
gymnast
halfback
halve
hamster
hoffman
hotdog
hulk
huxley
hyena
hypotheses
immigrant
inconvenient
inexact
informant
inhumane
input
interviews
israeli
istanbul
jacques
jitter
jujube
kafka
kamchatka
keyboard
kidnapping
kiwi
knowhow
kremlin
landcode
larynx
lincoln
lunchroom
luxe
macbeth
magtape
major
masker
maxwell
mazurka
megahertz
mysteries
native
newton
nihilist
object
ohm
onyx
optimum
oxford
paperback
papyrus
partner
persistent
pigment
pneumococcus
poet
popcorn
portfolio
potpourri
potsdam
projector
prospectus
quota
reflex
rembrandt
revue
rhesus
samovar
sandwich
scherzo
sheriffs
showman
shuttle
sightseeing
sleep
snob
society
software
squaw
stanza
stewards
stockholm
stopwatch
strychnine
studio
stuttgart
sweatshirt
symposium
tableau
teamwork
tokyo
tomahawk
tonic
transfer
trapezium
trekking
triplet
turf
turquoise
update
upgrade
vacuum
virgin
voltmeter
walrus
wonderland
workshop
wyoming
xylophone
yoga
yucca
zigzag
zwei

The list contains 1514 characters.

Digram Frequency Table for the NLUK-210 List.

	#	a	b	c	d	e	f	g	h	i&	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
#	-	21	21	19	14	10	7	5	9	9	3	7	5	8	3	5	13	1	4	21	11	2	3	4	1	2	2
a	14	1	3	9	10	1	2	4	2	2	1	1	11	6	28	1	6	1	13	4	8	4	-	3	2	1	2
b	1	10	1	-	1	3	-	-	1	2	1	-	2	-	-	9	-	-	6	1	-	5	-	-	-	2	-
c	1	3	1	2	1	2	-	-	12	3	-	8	1	-	-	15	-	1	2	1	8	3	-	-	-	1	1
d	10	4	1	1	1	16	1	1	1	7	1	-	1	1	1	6	-	-	1	1	1	1	1	1	-	1	-
e	29	5	2	6	3	3	1	1	1	3	1	1	7	2	18	1	1	2	24	6	9	2	2	3	7	2	1
f	1	1	1	-	-	3	3	1	-	1	1	1	3	1	-	5	-	-	1	1	1	2	-	-	-	-	-
g	6	3	-	-	-	4	-	1	2	1	-	1	1	1	1	1	-	-	2	1	1	2	-	1	-	1	1
h	3	9	1	-	-	8	-	-	-	3	-	-	1	1	1	7	-	-	1	1	1	4	-	1	-	3	-
i	5	3	1	2	2	6	1	4	1	-	1	1	3	2	24	2	3	1	2	13	4	5	1	1	2	-	1
j	-	1	-	-	-	3	-	-	-	1	-	-	-	-	-	3	-	-	-	-	-	5	-	-	-	-	-
k	9	5	-	-	1	2	1	-	1	3	-	1	1	-	1	1	1	-	1	1	1	1	-	-	-	1	-
l	5	7	1	1	1	8	1	1	1	10	-	1	4	1	1	4	1	-	1	1	1	3	1	1	-	1	-
m	12	13	1	1	-	6	1	-	-	3	-	-	1	1	1	2	1	-	-	1	-	1	-	1	-	1	-
n	16	7	1	7	8	13	1	6	1	5	1	2	1	1	1	3	1	1	1	2	18	1	1	1	1	2	1
o	4	2	3	7	5	1	2	2	1	1	1	2	6	5	11	1	7	1	19	3	6	4	1	5	1	1	1
p	5	3	-	1	1	6	1	1	1	3	-	-	1	1	1	9	2	-	2	1	1	1	-	1	-	2	-
q	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	11	-	-	-	-	-
r	16	14	1	1	7	9	1	1	1	11	-	3	1	1	3	5	1	1	1	1	10	2	1	1	-	2	1
s	16	3	1	3	1	4	1	-	6	6	1	1	1	1	1	2	1	1	1	2	20	1	-	1	-	1	-
t	40	9	-	1	1	17	1	1	2	5	-	1	1	1	1	9	1	-	6	3	3	5	-	1	-	1	2
u	1	3	3	2	2	7	1	1	1	2	1	-	2	10	5	2	3	1	5	8	5	1	1	-	2	1	-
v	-	2	-	-	-	3	-	-	-	3	-	-	-	-	-	2	-	-	1	-	-	1	-	-	-	-	-
w	2	6	1	-	-	4	-	-	1	4	-	1	1	1	1	3	-	-	1	1	1	-	-	-	-	1	-
x	5	1	-	1	-	1	1	-	-	1	-	-	1	-	-	1	1	-	-	-	1	1	-	1	-	1	-
y	7	1	1	2	1	1	-	1	-	-	-	-	1	2	1	3	1	-	1	2	1	1	-	1	1	-	-
z	2	3	-	-	-	2	-	-	-	2	-	-	-	-	-	1	-	-	-	-	-	1	-	1	-	1	1

Legend:

The ”#” code denotes a blank. A − denotes a zero count, and was used in this table instead of 0 because of its lower perceptual density

B Some basic statistics of the collected data

Analysis for 172 sets (210 words each)

Variable:	nstrok	npiece	ycorp	slant	width	nbars	ndots
Min	22.2	1.5	1.0	51.9	12.1	0.0	0.0
Max	35.8	9.2	5.2	110.8	36.6	1.0	1.2
GrandMean	27.2	5.2	2.1	83.8	22.2	0.1	0.5
SD	2.5	2.4	0.8	15.8	5.6	0.2	0.2

```
Legend:

nstrok     Average number of velocity-based strokes/word
npiece     Average number of pen-down segments/word
ycorp      Average vertical size of small letters (corpus, "x"-size) in [mm]
slant      Average angle of downstrokes at point of max. velocity  [degrees]
width      Average horizontal size of words in [mm]
nbars      Average number of vertical bar strokes/word
ndots      Average number of dots/word
```

C Writer Names

Typical Dutch names were attached to the writer sets, to be able to refer to the specific styles later.

D Coarse writing style classification on the basis of the average number of pen-down pieces per word

	writer	Npiece /word	standard deviation	sex	self-reported style
1	ineke	1.49	0.69	F	CURSIVE
2	angelien	1.56	0.74	F	CURSIVE
3	onno	1.60	0.72	M	CURSIVE
4	floris	1.79	0.85	M	CURSIVE
5	jeroen	1.86	0.93	M	CURSIVE
6	ruud	2.27	1.19	M	CURSIVE
7	johan	2.32	1.35	M	CURSIVE
8	willem	2.59	1.38	M	CURSIVE
9	gerrit	2.69	1.49	M	MIXED
10	koos	2.82	1.70	M	CURSIVE
11	miep	3.49	1.65	F	CURSIVE
12	piet	4.09	1.74	M	MIXED
13	loesje	4.65	1.72	F	MIXED
14	mark	4.91	1.82	M	MIXED
15	marieke	5.47	1.96	F	MIXED
16	heleen	5.58	1.93	F	CURSIVE
17	corrie	5.70	2.11	F	MIXED
18	juliana	6.10	2.01	F	MIXED
19	martijn	6.17	2.20	M	MIXED
20	hannie	6.30	1.99	F	MIXED
21	klaas	6.44	1.93	M	MIXED
22	janneke	6.55	2.31	F	MIXED
23	klaartje	6.58	2.18	F	MIXED
24	saskia	6.77	2.21	F	MIXED
25	katrien	6.96	2.21	F	MIXED
26	moniek	7.26	2.48	F	MIXED
27	kees	7.55	2.43	M	MIXED
28	eelco	7.60	2.06	M	MIXED
29	annemiek	8.00	2.36	F	MIXED
30	anton	8.05	2.20	M	MIXED
31	teun	8.22	2.48	M	PRINT
32	joost	8.32	2.39	M	PRINT
33	karel	8.60	2.51	M	PRINT
34	koen	8.80	2.61	M	MIXED
35	beatrijs	8.89	2.56	F	PRINT