Forensic Science International 126 (2002) 145–149

Forensic Science International

www.elsevier.com/locate/forsciint

A preliminary investigation into handwriting examination by multiple measurements of letters and spacing

Stephen Ling*

School of Applied Sciences, University of Glamorgan, Pontypridd, Mid Glamorgan, Wales CF37 1DL, UK

Received 11 May 2001; received in revised form 25 July 2001; accepted 13 February 2002

Abstract

Theoretically, the scientific measurement of letters, letter spacing and word spacing provides a mathematical basis for the statistical comparison of handwriting. Ten volunteers were asked to write the same short passage. Ten features were measured, up to 10 times each, to characterise each writer. Measurements of repeat handwriting 2 years later failed to identify the writer, and this was attributed to a change in the height of the middle zone. Measurements of 'it' and 'th' spacings were the least affected. © 2002 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Handwriting; Measurement; Discrimination; Characterise

1. Introduction

Computers are being used for handwriting examination. Found, Rogers and Schmidt [1] describe a computer program to automatically take measurements from scanned images. Houde [2] describes the grey scale equalisation, and Hicks [3] describes the use of Adobe 'Photoshop' to adjust the digital images and to present evidence. Here, the computer is used to make detailed measurements of letters and spacing.

In questioned document examination, it may be required to say if two hand-written documents are by the same author, or alternatively that they are definitely not by the same author. One document will be a questioned document and the other a known document. The questioned document could be a diary, an anonymous note or a will for example. The known document is handwriting by a known person provided under controlled conditions, or perhaps an undisputed personal letter.

Documents are examined to see how the letter and letter joins are formed and to observe the general pattern of the writing. The examiner then compares the questioned and known writing to reach a conclusion. To follow an explanation, it is hard for anyone to remember the pattern of writing and then to compare the remembered image to another. So it is helpful to select similar sections of text from the two writings, to photograph them and possibly enlarge them, and then to place them side by side. Fig. 1 shows the kind of comparison to demonstrate a similarity or a difference. Osborn [4] and Harrison [5] both describe this well-established method. However, the method uses a spot sample to demonstrate a generalisation. Within a passage of writing, a person might show several different constructions of the same letters or letter combinations. Eldridge et al. [6] identify those features that most reliably discriminate between writers. They conclude that the best features are:

- the top of staff closure in the letter 'k';
- the top of staff closure of 'f' in 'of' combinations;
- the top of staff closure of 'h' in 'th' combinations;
- the letter join in 'th' combinations.

Although arrived at by a numerical process, it is still not possible to use their conclusions to quantify the probability that two documents are by the same writer. Caution is needed when quantifying such a probability, which may not be justifiable if queried. To avoid this difficulty, Totty [7] quotes a scale of seven possible conclusions to express degrees of certainty without using numbers. They are, abbreviated, in order of decreasing likelihood that the documents are by the same writer:

- positive identification;
- strong probability identification;
- weak probability identification;

* Tel.: +44-1443-482-111; fax: +44-1443-482-285.
E-mail address: sjling@glam.ac.uk (S. Ling).

0379-0738/02/$ – see front matter
PII: S0379-0738(02)00048-8

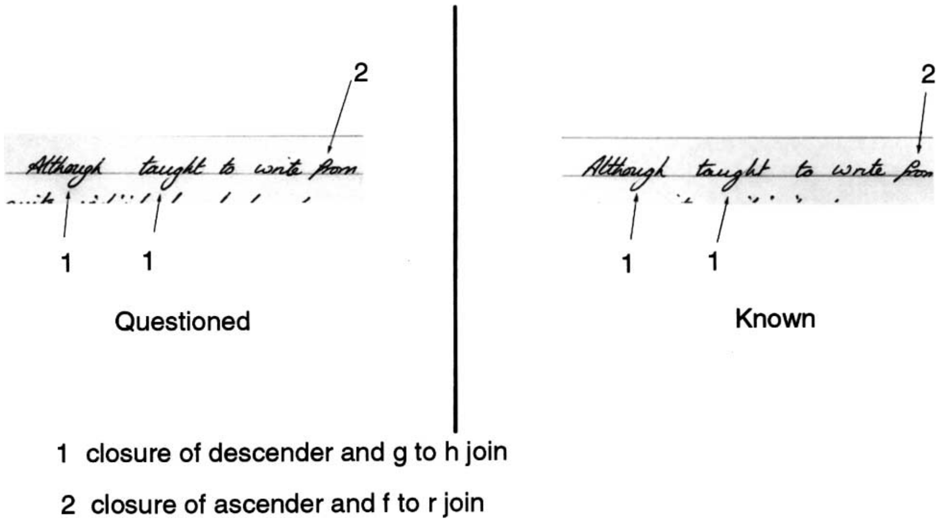

Fig. 1. A comparison of questioned and known handwriting.

- inconclusive;
- weak probability elimination;
- strong probability elimination;
- positive elimination.

Although widely used by others, it is my view that these conclusions lack objectivity. The description of weak probability is ineffective, and the strong probability options demand quantification. Particularly with the widespread use of DNA evidence, we are becoming accustomed to the quantification of probability to the extent that the word ‘probability’ without quantification may be viewed unscientific.

Taking measurements in document examination is a valid addition to the interpretation of letter shapes. Osborn [4] describes microscopes with which the necessary measurements can be made. He warns that ‘any measurements must be of discernible features if they are to convince a jury’. The numbers or measurements then add to the conclusion rather than forming the conclusion itself.

More recently, Buquet and Manchon [8] apply variance analysis to the measurement data, concluding that the technique is particularly useful for comparing Semitic scripts.

The objectives of this study were to test the discrimination between authors by some easily understood measurements of letters and word spacing. This was achieved through statistical comparison of the measurements.

2. Method

Ten volunteers were, simultaneously, asked to copy down a passage using the same A4-size ruled paper and any biro pen. The passage of writing was:

> Although taught to write from the same textbooks, our handwriting is quite individual. I do not profess to understand the mathematical possibilities, but the existence of at least five different ways to write each letter allows the author a huge permutation of letter profiles.

The passage was chosen to include ‘of’ and ‘th’ letter combinations as the words ‘of’ and ‘the’ and within longer words. Differences in letter combination designs were expected and this might contribute to the possibility of discrimination.

The samples of handwriting were scanned on a flatbed scanner and examined using Adobe ‘Photoshop’ software. The following measurements were taken:

1. Word spacing—the first 10 measurements, not including line endings or spaces containing punctuation marks, in mm. The shortest distance was taken.
2. The space between the ascenders of ‘th’ combinations—seven measurements, in mm. Measurements were made just above the middle zone. The middle zone is defined in Fig. 2.
3. The space between ‘i’ and ‘t’ in ‘it’ combinations—five measurements, in mm. Measurements were made just below the top of the middle zone.
4. The space between the ascending parts of the letter ‘u’—10 measurements, in mm.
5. The height of the letter ‘o’ at its greatest height—10 measurements, in mm.
6. The width of the letter ‘o’ at its widest point—10 measurements, in mm.

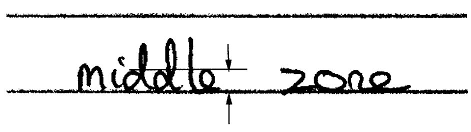

Fig. 2. The middle zone.

7. The slopes of the ascenders and descenders in letters ‘h’, ‘b’, ‘d’, and ‘p’—10 measurements, expressed as a gradient, positive if sloping forwards and negative if sloping backwards.
8. The height of ascenders—10 measurements, from the first 10 different words omitting capital letters, in mm.
9. The height of the middle zone as shown in Fig. 2—10 measurements from the first 10 different words, in mm.
10. The depth of descenders—10 measurements from the first 10 different words omitting capital letters, in mm.

Repeat measurements were taken from one sample of writing to estimate the accuracy of the measuring technique.

In a separate check, one volunteer’s handwriting was measured again 2 years later. The sample of handwriting was different but the person taking the measurements was the same, and the initial measurements were repeated to successfully confirm consistency of the procedure.

3. Results

3.1. Differences between volunteer writers

Table 1 shows the average values of the measurements numbered 1–10, referring to authors as writer A–J.

Repeat measurements indicated a measurement accuracy of 0.2 mm.

For each type of measurement, the averages spanned a three- or four-fold difference between writers. The standard deviations, given in Table 2, implied that if the distributions were normal, the measurement of several different features would be sufficient to characterise handwriting. One or other of the measurements would reveal a difference between the writers that was statistically significant.

With the exception of the slope, all measurements had to be positive, causing skewed distributions. A *t*-test of the significance of differences between averages is not appropriate for non-normal distributions. The Wilcoxon test, described by Walpole [9], is valid for any distribution. This test ranks observations and sums the ranks of each group. The test does not provide any information about the size of any difference between groups. It was used to test the null hypothesis that ‘the mean of the measurements of writer E could be the same as the mean of the measurements of the writer in question.’ A similar comparison is possible with each of the other writers. Table 3 shows the results of comparing each of the others with writer E. This demonstrates that the writing was characterised by the complete set of measurements. Differences between each of the writers and writer E could be detected.

Table 1
Mean values of measurements 1–10 from handwriting by authors A–J

	A	B	C	D	E	F	G	H	I	J
1	4.68	4.42	9.96	4.88	6.04	7.66	5.84	5.44	6.8	2.16
2	1.46	2.66	1.37	1.54	1.54	1.43	3.66	0.97	2.29	2.17
3	1.68	3.4	2	2.72	1.88	1.72	2.36	1.04	2.44	1.96
4	1.5	3.22	1.9	2.54	1.8	3.78	2.73	1.1	3.66	1.9
5	1.5	2.92	1.66	3.3	1.24	2.96	2.86	1.42	2.54	1.64
6	1.28	3.28	1.68	3.28	1.62	4.22	3.34	1.14	2.98	1.6
7	0.91	−0.15	−0.25	0.34	−0.16	0	0.16	0.33	0.03	0.44
8	1.86	1.4	2.62	1.5	2.14	3.76	2.82	3.02	1.12	0.7
9	1.7	2.98	1.84	3.14	1.3	3.08	2.86	1.46	2.6	1.64
10	2.12	3.06	2.29	2.58	2.04	4.7	2.78	2.2	2.14	1.42

Table 2
Standard deviation of measurements in Table 1

	A	B	C	D	E	F	G	H	I	J
1	1.53	1.25	1.54	70.64	2.02	1.48	1.15	1.38	1.59	0.62.
2	0.43	0.72	0.31	0.25	0.22	0.65	0.85	0.21	0.34	0.27
3	0.33	0.58	0.32	0.30	0.41	0.54	0.71	0.22	0.38	0.17
4	0.17	0.45	0.32	0.25	0.35	0.74	0.48	0.30	0.46	0.19
5	0.17	0.33	0.19	0.27	0.25	0.48	0.25	0.26	0.27	0.16
6	0.25	0.57	0.27	0.33	0.3	0.57	0.42	0.34	0.24	0.23
7	0.13	0.14	0.06	0.13	0.11	0.0	0.19	0.06	0.09	0.11
8	0.4	0.48	0.38	0.55	0.42	0.32	1.1	0.94	0.53	0.19
9	0.22	0.29	0.23	0.31	0.25	0.25	0.31	0.37	0.3	0.13
10	0.79	0.82	1.17	0.95	0.4	1.81	0.99	0.71	0.62	0.26

Table 3
Significance of differences between author E and the other authors using the Wilcoxon 2% level of significance

	A	B	C	D	E	F	G	H	I	J
1	No	No	Yes	No		No	No	No	No	Yes
2	No	Yes	No	No		No	Yes	Yes	Yes	Yes
3	No	Yes	No	Yes		No	No	Yes	No	No
4	No	Yes	No	Yes		Yes	Yes	Yes	Yes	No
5	Yes	Yes	Yes	Yes		Yes	Yes	No	Yes	Yes
6	No	Yes	No	Yes		Yes	Yes	Yes	Yes	No
7	Yes	No	No	Yes		Yes	Yes	Yes	Yes	Yes
8	No	Yes	No	Yes		Yes	No	Yes	Yes	Yes
9	Yes	Yes	Yes	Yes		Yes	Yes	No	Yes	Yes
10	No	Yes	No	No		Yes	No	No	No	Yes

The height of the middle zone, the height of the letter 'o' and the slope were the most discriminating measurements, and the spaces between words was the least discriminating. By inspection of Table 3 it may be suggested that some measurements had a similar effect, or pattern of discrimination, and therefore, that they might not be independent. For example:

(a) the height of the middle zone and the height of the letter 'o';
(b) the width of the letter 'o', the spacing of 'th' and 'it' combinations, and the spacing of the ascenders in the letter 'u';
(c) the height of the ascender and depth of the descender.

In contrast, both the slope and word spacing may be independent of the other measurements.

3.2. Differences between writer E and writer E after 2 years

Measurements from a sample of writer E's handwriting 2 years later are shown in Table 4. The height of the middle zone, as defined in Fig. 2, and also the height of the letter 'o' were significantly different 2 years later.

Table 4
Repeat measurements of writer E after 2 years

	Average initial measurement	Average final measurement	Significance of difference
1	6.04	5.48	No
2	1.54	1.34	No
3	1.88	1.76	No
4	1.8	1.68	No
5	1.24	1.66	Yes
6	1.62	1.88	No
7	−0.16	−0.04	No
8	2.14	2.44	No
9	1.3	2.04	Yes
10	2.04	3.14	No

4. Discussion

4.1. Discussion of differences between writers

Eldridge et al. [6] suggest the need for a way to rank the discriminating power of a feature and propose a 'discrimination index'. In their experiments they found that the construction of the top of staff in 'of' and 'th' combinations, and the letter 'k' had high discrimination indices. The ability to reject the null hypothesis more times in these experiments is equivalent to a high discrimination index.

In these experiments, most discrimination was achieved when:

(a) One or other of the writers being compared has an average measurement that is unusual. It is that the average value is close to the limit of the range normally observed.
(b) The variation within the sample is small, making the differences between the samples more significant. This is most likely when the writing is 'tidy.'
(c) The variation between samples is large, obviously.

No single type of measurement has the potential to discriminate in all cases. Therefore, the results show the need to make several different types of measurement. In order to be most useful, each different type of measurement must be independent. As an initial suggestion, the spacing of 'th' combinations, measurements of the height of ascender, and slope appear to be independent. However, a further complexity is discussed below.

Osborn [4] recommends that measurements must be of discernible features to be convincing. Fig. 3 shows some comparisons from which we might argue that:

(a) The space between the ascenders of 'th' combinations is greater for writer B than for writer E.
(b) Writer B's ascenders are smaller than writer E's.
(c) Writer B's middle zones are bigger than those of writer E.

In Fig. 3, writer B joins 't' to 'h' in 'th' letter combinations by the crossbar, except for the second example where

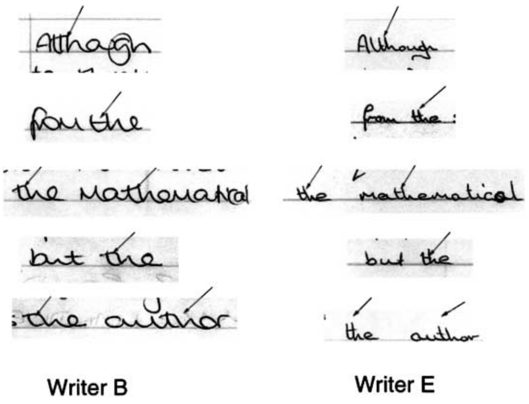

Fig. 3. Some differences between the handwriting of writer B and writer E.

the bottom of staff is used. Writer E uses the bottom of staff in all five examples. Is it possible that the different constructions contribute to the significantly different measurements of the space between the ascenders of 'th' letter combinations (measurement 2) between writer B and writer E? If so, measurement 2 provides an alternative way to quantify the difference in construction—the original way being that of Eldridge et al. [6], which is to count the frequency with which each construction is used.

In the example shown in Fig. 3 the way in which 't' to 'h' joins are made could contribute to the observed difference in the height of ascenders (measurement 8). Therefore, it may be suggested that measurement 2 and measurement 8, having the same root cause, are not independent.

This preliminary study suggests the need for further investigation to establish which aspects of handwriting are independent, and which aspects remain constant over time. It was found that writing can be more easily identified if it is tidy, and this raises the suggestion that tidiness itself is a separate identifying feature, which can be measured as a lack of variability.

4.2. Discussion of differences between writer E and writer E after 2 years

The difference between writer E and writer E after 2 years is mainly that the middle zone was bigger in the later writing. This was not due to a change in the spacing of the lined paper, which was identical. Further, since many of the measurements were smaller in the later writing, it is not possible to correct them by applying a scale factor.

It is not possible to say if this variation lies within a day to day variation because only the two samples of writing were considered. Further measurements under controlled conditions are required to establish how much variation occurs on a day to day basis.

5. Conclusions

Simple measurements discriminated between writers, but could not identify another sample of handwriting by the same writer 2 years later. In this study the writer used a bigger middle zone in the later handwriting.

Multiple measurements allow statistics to be applied to handwriting examination, but a significant difference in measurements is not sufficient to prove that the writing is by a different person.

A simple method of measurements will not easily equal the capacity for complexity of the trained eye to recognise an individual's handwriting.

References

[1] B. Found, D. Rogers, R. Schmittat, A computer program to compare the spatial elements of handwriting, Forensic Sci. Int. 68 (1994) 195–203.

[2] J. Houde, Image enhancement for document examiners using the personal computer, J. Forensic Sci., JFSCA 38 (1) (1993) 143–146.

[3] A.F. Hicks, Computer imaging for questioned document examiners. Part 1. The benefits, J. Forensic Sci., JFSCA 40 (6) (1995) 1045–1051.

[4] A.S. Osborn, Questioned Documents, 2nd Edition, Boyd, New York, 1929.

[5] W.R. Harrison, Suspect Documents, their Scientific Examination, Praeger, New York, 1958.

[6] M.A. Eldridge, I. Nimmo-Smith, A.M. Wing, R.N. Totty, The variability of selected features in cursive handwriting: categorical measures, J. Forensic Sci. Soc. 24 (1984) 179–219.

[7] R.N. Totty. Recent Developments in Handwriting Examination. Forensic Science Progress, Vol. 5, Springer, Berlin, 1991, pp. 91–128.

[8] A. Buquet, P. Manchon, Handwriting examination-appliance of variance analysis to measurement data: Semitic handwriting, Int. Criminal Police Rev. 414 (1988) 35–39.

[9] R.E. Walpole, Introduction to Statistics, 2nd Edition, Macmillan, New York, 1974.