AUTHENTICATION BY KEYSTROKE TIMING: SOME PRELIMINARY RESULTS

PREPARED UNDER A GRANT FROM THE NATIONAL SCIENCE FOUNDATION

R. STOCKTON GAINES, WILLIAM LISOWSKI,
S. JAMES PRESS, NORMAN SHAPIRO

R-2526-NSF
MAY 1980

Rand
SANTA MONICA, CA. 90406

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **MAY 1980**	2. REPORT TYPE	3. DATES COVERED **00-00-1980 to 00-00-1980**
4. TITLE AND SUBTITLE **Authentication by Keystroke Timing: Some Preliminary Results**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Rand Corporation,1776 Main Street,PO Box 2138,Santa Monica,CA,90407-2138**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**51**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

The research in this report is supported by the National Science Foundation under Grant No. MCS76-00720.

Published by The Rand Corporation

AUTHENTICATION BY KEYSTROKE TIMING: SOME PRELIMINARY RESULTS

PREPARED UNDER A GRANT FROM THE NATIONAL SCIENCE FOUNDATION

R. STOCKTON GAINES, WILLIAM LISOWSKI,
S. JAMES PRESS, NORMAN SHAPIRO

R-2526-NSF
MAY 1980

Rand
SANTA MONICA, CA. 90406

PREFACE

This report was prepared as part of Rand's research project on Computer Security, sponsored by the National Science Foundation under Grant No. MCS76-00720.

The growing use of computers to store sensitive, private, and classified information makes it increasingly important to be able to determine with a very high degree of confidence the identity of an individual seeking access to the computer. This report summarizes preliminary efforts to establish whether an individual can be identified by the statistical characteristics of his or her typing.

The investigation was carried out under the joint direction of Stockton Gaines and Norman Shapiro, who are responsible for the central idea of using keystroke timing as the basis for an authentication system. They also developed the textual material upon which the experiment was based, and they conducted the experiment. James Press developed the statistical model for authentication, directed the analysis of the experimental data, and drafted the report. William Lisowski programmed the authentication procedure for the computer, developed programs for analyzing the data, and ran the data through the routines.

SUMMARY

Can people be identified by the way they type? To investigate this question, an experiment was carried out at Rand, in which seven professional typists were each given a paragraph of prose to type, and the times between successive keystrokes were recorded. This procedure was repeated four months later with the same typists and the same paragraph of prose. By examining the probability distributions of the times each typist required to type certain pairs of successively typed letters (digraphs), we found that of the large number of digraphs represented in most ordinary paragraphs, there were five which, considered together, could serve as a basis for distinguishing among the subjects. The implications of this finding are that touch typists appear to have a typing "signature," and that this method of distinguishing subjects might provide the basis for a computer authentication system.

CONTENTS

I. INTRODUCTION

This report describes the preliminary results of an investigation of the feasibility of using keystroke timing as the basis for authenticating individuals seeking access to sensitive information stored in a computer. In many such applications, authentication might be carried out using software stored in the computer itself. The fundamental question that must be answered is, Do people type in timing patterns that are so individual that one typist can be distinguished from another, with extremely high reliability, on the basis of their typing "signatures"?

There is some a priori reason to believe that individuals type differently in a statistically significant way. For instance, it has been known that people who use a telegraph key develop a distinctive "fist" or telegraphic style that can be recognized. Amateur radio operators can often tell which of their friends is transmitting, before direct identification is received. Moreover, it has been discovered that not only is the *form* of an individual's written signature unique and distinctive, so are other aspects of writing a signature. The pen pressure used in producing the signature and the acceleration of the pen are variables that can be measured and whose patterns can be associated very accurately with the signer. Because the act of typing is mainly one of involuntary control of finger movements, at least in the case of a skilled typist, we had reason to hope at the beginning of this investigation that typing patterns would be both different enough between individuals and consistent enough over time that authentication based on the timing characteristics of typing would be feasible.*

To investigate the extent to which typing signatures exist, and to evaluate whether or not individuals can actually be authenticated on the basis of them, we designed an experiment involving a typing

---

*We also examined the earlier efforts to analyze individual typing behavior reported by Coover (1923), Dvorak et al. (1936), Harding (1933), Lahy (1924), Neal (1977), Ostry (1977), and Rochester et al. (1967).

"test," which we administered to subjects. After analyzing the statistical properties of the subjects' typing patterns, we developed a statistical model for authenticating subjects; we then applied the model to the data from the experiment. The results were sufficiently promising to suggest both that more extensive experimentation should be undertaken and that the development of the statistical model should be broadened to extend its applicability.

The experiment is described in Sec. II. Briefly, it involved the collection of samples of keystroke timing from seven individuals at two different times, separated by four months. However, only six were available for the second data collection. The statistical model used to analyze the data is described in Sec. III, and the detailed analysis of those data is presented in Sec. IV. The mathematical details of the model are given in the Appendix.

Prior to performing our detailed analysis, we conducted the following informal experiment: One member of the project staff was given all the data, with the names of the individuals removed. The data consisted of each individual's average time for typing each digraph, i.e., each pair of letters typed successively in a text.

This person then tried to match the data from the first period with those from the second period on an individual-by-individual basis. He was able to do this with 100 percent success; he was even able to identify the set of data from the individual who took the test the first time but was not present for the second session. The comparison was simply performed by eye, without using any sort of formal analysis routines. This result considerably strengthened our hypothesis that individual typing characteristics are substantially different between individuals.

There are, of course, many ways in which a "signature" might occur in an individual's typing patterns. We might have looked, for example, at the time to type entire words, entire sentences, or entire paragraphs. However, we chose to examine digraphs, because they seemed the most elemental typing units. Future analyses might explore the potential of using other data for authentication. The success we achieved with digraphs strengthened our belief that they are useful for authentication, but we have by no means ruled out the possibility that other measures might be even more useful.

II. THE EXPERIMENT

Our experiment on keystroke timing involved having six touch typists (professional secretaries at Rand) type each of three specially prepared texts.* They were then asked to repeat this task four months later, using precisely the same texts. We were thus able to study variations across people who took the same test at the same time, and we could also study typing consistency for a given individual typing the same text at a later time. Two of the six typists studied were left-handed and four were right-handed.

The three texts are reproduced in Figs. 1 through 3. The first (Text 1) was designed to read as ordinary English text; the second (Text 2) is a collection of "random" English words; and the third (Text 3) is a collection of "random" phrases. We originally hoped to be able to make separate conclusions about how individuals differ in their typing of the three kinds of textual material. As it turned out, however, there was insufficient information in any one of the texts to permit statistical inferences to be drawn from that text alone. Therefore, we pooled the information in the three texts, so our data base was developed by using the three texts as if they were one long continuous text.

The typing keyboards were part of a PDP-11/45 computer system. A timer was installed within the system to record the time at which each key was struck. A small program then calculated the time between each pair of successive letters, or digraphs. The time between successive letters is referred to as the "digraph time." Thus, the time it takes to type *io* is one digraph time, and the time to type *on* is another. (Although we have so far analyzed only digraph times, we can envision using trigraphs such as *ion* or tetragraphs such as *tion*, as well.) The digraphs we have considered involve only lower-case letters and spaces; upper-case letters, carriage returns, punctuation, and

*There were originally seven subjects, but one was not available to complete the experiment.

Most Americans now do at least some of their buying on credit and most have some form of life, health, property or liability insurance. Institutionalized medical care is almost universally available. Government social services programs now reach deep into the population along with government licensing of occupations and professions, federal taxation of individuals, and government regulation of business and labor union affairs. Today government regulates and supports large areas of economic and social life through some of the nation's largest bureaucratic organizations, many of which deal directly with individuals. In fact, many of the private sector record keeping relationships discussed in this report are to varying degrees replicated in programs administered or funded by federal agencies.

A significant consequence of this marked change in the variety and concentration of institutional relationships with individuals is that record keeping about individuals now covers almost everyone and influences everyone's life, from the business executive applying for a personal loan to the school teacher applying for a national credit card, from the riveter seeking check guarantee privileges from the local bank to the young married couple trying to finance furniture for their first home. All will have their creditworthiness evaluated on the basis of recorded information in the files of one or more organizations. So also with insurance, medical care, employment, education, and social services. Each of those relationships requires the individual to divulge information about himself, and usually leads to some evaluation of him based on information about him that some other record keeper has compiled.

Fig. 1 — Sample 1

plasma wring fork gnome twitch vapor proms doze half blur whimper
fib fuzzy eggnog docent wry placard gyp pablum duffle twenty
extract wheeze ward churn endurable bystander legible avid razz
vivisect swat hull smirk paams type active keys lyse skirmish
frenzy fox extra hubby swamp excite skies keg stanza pun kill
form sweaty foxy half smuggler lava excise under duffer fuzzy
active churn smirk half form exise twitch under docent legible
extract wheeze ward pablum wring doze smuggler keys skirmish
bystander gnome endurable swamp plasma vapor avid half frenzy
stanza placard prams vivisect keg fork gyp sweaty pun skies blur
eggnog razz type swat lyse hubby excite kill duffle foxy lava wry
fib proms hull fox extra twenty whimper duffer pun form ward
churcn fork eggnog plasma skirmish endurable razz active foxy swat
excite vivisect twenty placard fuzzy wheeze fox smuggler avid
hull fib type docent bystander prams blur pablum doze lyse
extract duffer keys vapor duffle under skies wry whimper swamp
kill smirk twitch keg frenzy sweaty hubby excise stanza gyp half
proms lava gnome wring half legible extra keys frenzy extract
swamp kill smuggler wring gyp plasma bystander vivisect half
active under wheeze stanza skies hubby placard type fuzzy
endurable legible duffer twenty doze skirmish pablum docent foxy
vapor ward blur eggnog pun proms fox excite lyse half twitch
duffle lava sweaty form avid prams smirk fork whimper keg gnome
hull extra churn excise wry swat fib razz eggnog duffer half
excite pun type placard bystander smuggler hull endurable frenzy
half keys skies legible hubby fork fib blur twitch swat skirmish
swamp wheeze gnome active gyp razz lyse extract duffle ward smirk
whimper excise prams avid proms wry fuzzy stanza vapor under doze
form pablum twenty docent lava plasma vivisect wring sweaty foxy
churn extra kill fox keg

Fig. 2 — Sample 2

This typing exercise is a strange jumble of awkward phrases, representing the quintessence of exquisite digraphs dictated by a foreign midget. It is a plethora of puzzling words under the guise of psychological authentication policy, although you may perceive it quizzically as an ambiguous wasteful plot to overcome summertime melancholy. Your vituperations against this phenomenon will add to the dense psychodrama in which this impossible business is entwined. The hyphenated rhythms of this ridiculous nightmare may elicit smothered teardrops as well as excited little laughs. The psychotic excesses may lead to indefinite suspense or mumbling traditional sayings, or may just produce a kind of loud ringing in the ether. Whatever the consequence, enough mystic bifurcations dangled and untried will decimate the ranks of all but the most adventurous or mercenary. If it is rough, pound it; if lousy, fight it. All is fair in cybernetic war if plotted smartly.

The English jury snapped under the known betrayal, but still sent the European ragamuffin to the penitentiary. The earthenware was made from black milk, pounded to a chalky consistency. The sedentary safecracker succeeded by using a lubricated blue pencil. He would swear that a snafu was unsynchronized, although fencing at a high altitude was crass. The excluded sex rarely chuckles unless judiciously engaged in schoolwork. The phenomenal pansies growing aside the softball mound were fortuitous twins. Stubble in bulk should be checked. The suspected lubber did not wear sable onto the frigate. A blank lethargy results from a lackadaisical twiddling. If you are dumbfounded, you may quit and go dancing or bicycle on the promenade.

Fig. 3 — Sample 3

special characters have been ignored because of their relative infrequency in typewritten material. The digraph times ranged from a minimum of about 75 milliseconds to a maximum of several seconds (times were recorded to an accuracy of within 1 millisecond). The extremely high values probably represented some external interruption of the typing task. The typical digraph time was around 125 milliseconds.

Once we started analyzing the digraph times, it became clear that in future experiments we could avoid certain problems by building into our experimental texts a certain minimum number of replications of "important" digraphs; moreover, we would try to make the texts used in multiple-text experiments more unlike one another than those used in the initial experiment. Finally, we would use a larger number of subjects in subsequent experiments.

III. THE STATISTICAL MODEL

INTRODUCTION

The statistical model we have adopted assumes that a person (called the "originator") who will later desire to gain access to a computer types some predesignated text into the computer, which then retains information regarding the keystroke-typing time. Later, another person (called the "claimant") who wishes access to the computer and who makes a claim to being the originator is asked by the computer to type in another predesignated text. The computer must now compare the keystroke-typing time patterns of the claimant with those of the originator. If the two are the same, at least in terms of their statistical characteristics, then a system based upon our model will authenticate the claimant as being the same person as the originator; if the patterns do not match, the system will not authenticate the claimant and will not allow him to log on.

An authentication system can make two types of error: a "primary" error, in which an unauthorized person (impostor) is granted access to the computer; and a "secondary" error, in which the system fails to give access to an authorized person. While the terms "primary" and "secondary" are of course arbitrary, a primary error would, in most contexts, be much worse than a secondary error. (An exception would be the case in which a decisionmaker, such as an army general, must issue counterattack commands immediately, in response to an attack, and he must do it through a computer. If the computer security system fails to authenticate him and denies him access, precious minutes are lost while the general tries to get his counterattack started.)

The hypothetical authentication system considered here is based upon a statistical model that uses the classical theory of hypothesis testing. The basic ideas behind classical hypothesis testing have been amply described elsewhere, so they will not be repeated here. We will build and draw upon them, however.

In the authentication problem, we will use H to denote the hypothesis that the claimant and the originator are the same person, and A will denote the hypothesis that the claimant and the originator are different persons, i.e., that the claimant is an impostor.

We test H versus A in terms of the significance level of the test, which is normally written,

$$\alpha = P\{\text{rej. } H|H\} = P\{\text{making a secondary error}\}.$$

The probability of making the other kind of error is given by

$$\beta = P\{\text{rej. } A|A\} = P\{\text{making a primary error}\}.$$

In many problems of inference, α is taken to be .01, 05, or .10. We will also work in this range.

Ideally, we should attempt to simultaneously minimize α and β, but unfortunately, we cannot reduce one without increasing the other. In keeping with normal statistical practice, therefore, we will fix α in advance at some tolerably low level and try to keep β as small as possible.

In our problem, we will use a test statistic U that reflects the difference in keystroke patterns between the originator and the claimant. If the two individuals are the same person, U should be small (i.e., not significant, reflecting only random sampling variation), and we should not want to reject H. Therefore, the p-value corresponding to an observed U should be large ($\geq$ .05). If in fact the p-value is small, we generate a secondary error.

Alternatively, suppose the originator and the claimant are different persons. In this case, U should be large (significant), and we should want to reject H. Therefore, the p-value should be small (< .05). If in fact the p-value is large, we generate a primary error. These concepts are summarized in Table 1.

We derived the test procedure for our problem on the basis of a classical likelihood ratio test. The procedure is summarized below; the technical details of the derivation are given in the Appendix.

Table 1

ERROR CONCEPTS

	p-value Large (accept H)	p-value Small (reject H)
H is true: originator and claimant are the same person	No error	Secondary error
H is false: originator and claimant are different persons	Primary error	No error

AUTHENTICATION EQUATIONS AND PROCEDURE

The subject whose keystroke typing patterns are being evaluated (either claimant or originator) is asked to type a paragraph of prose, and the computer records the time between all successive keystrokes. For a judiciously selected group of digraphs, the authentication procedure will compare the digraph times from the claimant's sample with those from the originator's sample.

For example, the originator types the digraph *th* ten times in some nonrepetitive, prose context (to avoid "learning"), with a mean digraph time of 85 milliseconds and a standard deviation of 5 milliseconds. The claimant then types the *th* digraph 15 times, with a mean digraph time of 150 milliseconds and a standard deviation of 10 milliseconds. In this case, it seems likely that the claimant is an impostor.

The raw data collected in any real situation are likely to show that digraph times for a specific digraph are roughly log-normally distributed (see Sec. IV). Thus, their logarithms are approximately normally distributed. We assume in the authentication equations that the variables created by transformation from the raw data are approximately normally distributed.

We work simultaneously with r distinct digraphs, each of which is assumed to be typed M times by the originator and N times by the claimant. In fact, because of typing errors, subjects tended to type different numbers of replications of a given digraph. For example, if one typist inadvertently omitted a word that included a *th*, while all of the others made no errors involving a *th*, that typist would have one fewer replication for *th* than the others. For purposes of analyzing the text obtained from an originator, we selected the first M replications of a given digraph in the text (for the claimant, we selected the first N), and we ignored the remainder. M and N were determined as the smallest number of replications that occurred for all r digraphs. Thus, if there were three digraphs to be considered for the originator and one was replicated 12 times, another 15 times, and the third 15 times, we would select M=12, because there were at least 12 replications in all three (and the statistical model requires an equal number for all digraphs). We would then select for analysis the first 12 occurrences of each of the three types of digraphs in the originator's text.

We assume that the M+N digraph times for each of the r digraphs (that is, (M+N)r distinct times) are mutually independent. We know, of course, that this assumption is not strictly true, but we adopt it for simplicity as a first approximation to see if a system can eventually be developed around it. Clearly, the third time a *th* is typed in no way influences (or is influenced by) the fourth time a *th* is typed by the same person; nor is there generally any natural way to pair the digraph times for any particular pair of digraphs (identical or not).

We assume that the distribution of the time required to type a particular digraph has, after transformation to normality, the same variance for both originator and claimant. That is, the variance of the transformed digraph time distribution for a *th* will be taken to be the same for both originator and claimant, although variances for different digraphs such as *th* and *he* are permitted. The mean digraph times are of course permitted to differ from one another, both across digraphs and between claimant and originator; in fact, the test of hypotheses H versus A will be carried out on the basis of how the mean digraph times of claimant and originator compare. The assumption of equal variances for claimant and originator made above is justifiable on the basis of the well-known metatheorem in statistical theory:

Tests for equality of means, under normality, are fairly insensitive to violations of the assumption of equal variances. This is a robustness property of Student t-tests. Thus, the statistical model for authentication basically involves testing the hypotheses that the mean vectors (vectors of mean digraph times) for two multivariate normal populations are or are not the same, assuming that the two populations have the same diagonal covariance matrix (diagonal, because the digraph times are assumed to be independent).* A likelihood ratio test is carried out to develop an appropriate test statistic, and it is found, not surprisingly, that the test statistic is a function of the corresponding Student t-statistics for each of the digraphs. In fact, the test consists of adding 1 to the Student t-statistic for each digraph, then multiplying all of them together. A monotone function of this product is tested for significance.

EXTENSIONS OF THE MODEL

Extensions of this statistical model could conceivably involve development of models that permit different numbers of replications for different digraphs, unequal variances for the distributions of digraph times for claimant and originator, correlations of times for distinct digraphs, and perhaps a better approximation to the distribution of a product of independent beta variates than the one developed in the Appendix. Such extensions could increase the flexibility of an eventual authentication system and might improve the precision of such a system by providing statistical tests that are more powerful and make fewer errors. We might also develop a measure of sensitivity of the authentication tests based upon the notion of "power" of a test of hypotheses. We are considering an alternative model in which the parameters of the originator's digraph distributions are assumed to be

---

*It is clearly important to test this assumption. A fundamental problem, however, is that there is no natural pairing of digraphs that will permit us to compute the sample correlation of digraph times across N pairs. Alternatively, we computed sample correlations across the first occurring sets of pairs for a great many digraphs. In all such cases, the correlations were not significant at the 5 percent level of significance.

known, because the schema we envision should permit us to obtain large numbers of replications of digraphs for the originator, although probably not for the claimant.

IV. STATISTICAL ANALYSIS

BACKGROUND

The data collected in the experiment include typescripts of three structured texts, typed on two different occasions by six touch typists. The dates on which the experiment was administered, August 16 and December 14, 1977, were four months apart. Six subjects participated in the experiment, but all of them did not type all three texts each time. Typist 2 failed to type Text 3 in December; Typist 3 failed to type Texts 2 and 3 in August; and Typist 6 failed to type Text 2 in August. The missing data are summarized in Table 2.

Table 2

MISSING EXPERIMENTAL DATA

		August Session			December Session		
Typist	Typist's Handedness	Text 1	Text 2	Text 3	Text 1	Text 2	Text 3
1	Left	--	--	--	--	--	--
2	Right	--	--	--	--	--	X
3	Right	--	X	X	--	--	--
4	Left	--	--	--	--	--	--
5	Right	--	--	--	--	--	--
6	Right	--	X	--	--	--	--

The times for all digraphs in each of the texts were recorded at both sessions. The first question we addressed in the analysis of the data was, What is the distribution of digraph times for a given subject, for a given digraph, both in August and in December?

We began by developing computer plots of the histograms associated with each case. A sample of the histogram plots is given in Fig. 4. Each histogram is labeled with four codes: The first code indicates the number of the subject (1-6), the number of the text typed (1-3), and whether the test was taken in August (1) or in December (2). The second code is the digraph. Those entries that include a dash (such as

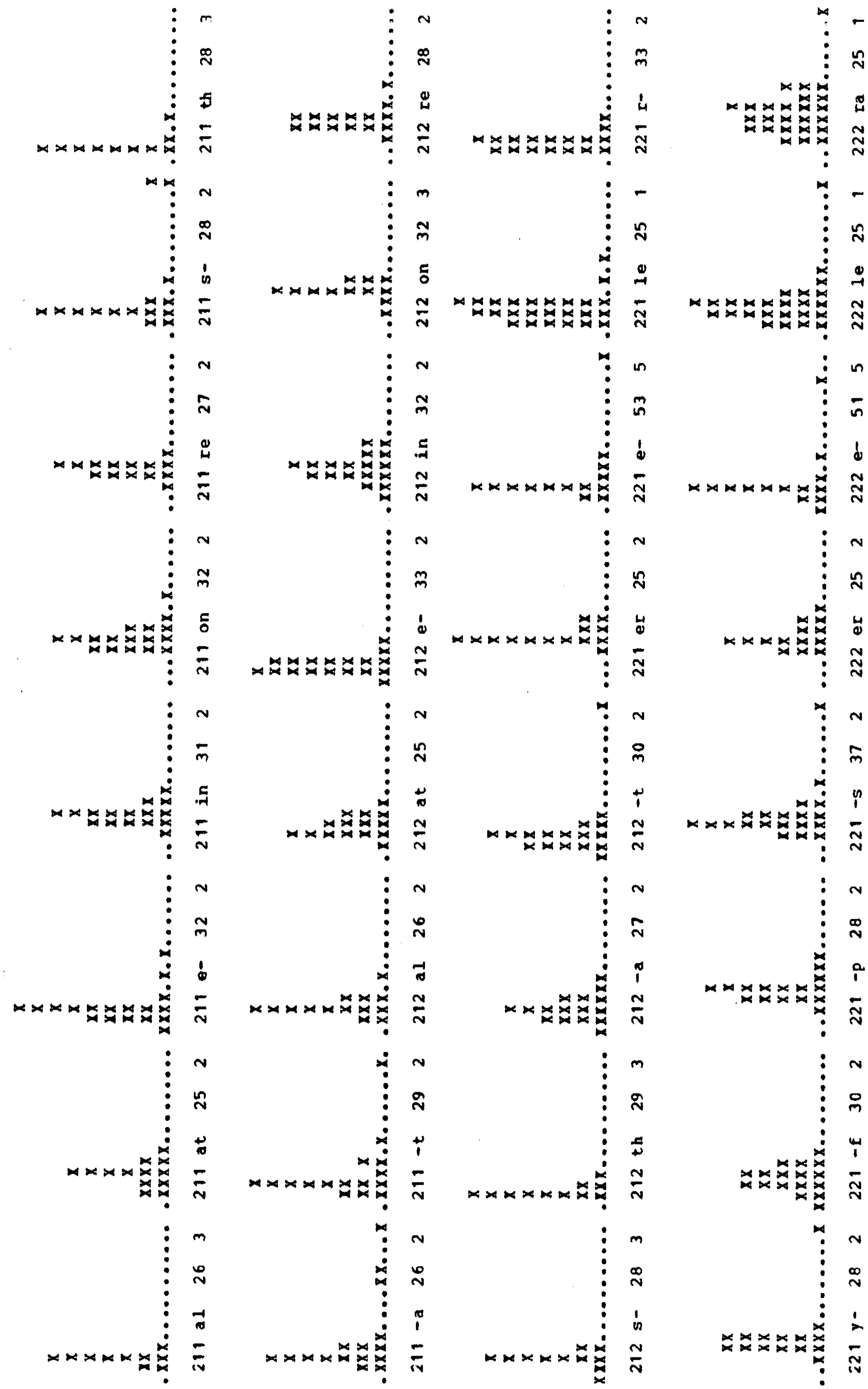

Fig. 4 —— Histogram plots of digraph times

e- or -*a*) indicate that the letter shown may be paired with any other character. The third code represents the number of replications of the digraph plotted in the histogram, and the fourth entry gives the scaling for the vertical scale in each plot (for example, a 3 indicates that each X in the histogram represents three replications, except perhaps for the topmost X in each column, which may represent one, two, or three replications). Thus, the first histogram in Fig. 4 represents the performance of Typist 2 on Text 1 in August; 26 replications of the digraph *al* are plotted, with each X representing three replications. The histogram on the second line labeled "212 *e*- 33 2" indicates that an *e* was followed by some other character 33 times in the sample.

The horizontal scale shows digraph time, measured in 25-millisecond intervals, starting with 50 milliseconds. Thus, the first column in each histogram shows the number of times the given digraph was typed in 50 to 74 milliseconds; the next column is for 75 to 99 milliseconds; etc. The rightmost column indicates digraph times of 400 milliseconds and above.

We hypothesized that the large tail in the distribution was caused by the typist sneezing, pausing, or whatever, while typing some digraphs. Accordingly, we removed all digraph times exceeding 500 milliseconds from the data, then reexamined the histograms. We still found long tails in the distribution, so we took the logarithm of all digraph times* and replotted the histograms in terms of the logged data (excluding the digraph times exceeding 500 milliseconds). These histograms tended to look much more normally distributed than any of the previous plots (although this was not true in all cases). The data obtained by removing the outlying digraph times and taking logs of all remaining observations will hereafter be referred to as the transformed data.

Now that the transformed data at least "looked" normally distributed, we proceeded to check further into how far the distributions

---

*The log transformation is a special case of the more general class of so-called Box-Cox transformations, used to induce normality of the transformed data (the more general class also includes power transformations). We decided, however, to ignore the possibility of achieving even better fits to normality with such transformations because of the exploratory nature of the analysis.

actually deviated from normality. We computed the first four sample moments of each set of the transformed data and then evaluated the mean, variance, skewness, and kurtosis. The skewness is defined as the third central moment divided by the variance raised to the power 3/2. The kurtosis is computed by subtracting 3 from the fourth central moment divided by the squared variance. Both the skewness and the kurtosis are zero in a normal distribution.

An illustrative collection of sample moments is shown in Fig. 5, for the case of Typist 6 (recall that 612 denotes Typist 6 typing Text 1 in December); n denotes the number of replications of each digraph. Inference regarding the population values was carried out as follows: Let

$$\theta \equiv \text{skewness} = \mu_3/\mu_2^{3/2}$$

$$\phi \equiv \text{kurtosis} = (\mu_4/\mu_2^2) - 3 \ ,$$

where

$$\mu_k \equiv E(X - EX)^k, \quad k = 2,3,4 \ ,$$

and E denotes expected value of a random variable.

It has been shown (see Cramér, 1946) that for large sample sizes, n, assuming X is normally distributed, it is approximately true that

$$\hat{\theta} \sim N(0,\tau_1^2), \qquad \hat{\phi} \sim N(0,\tau_2^2) \ ,$$

where

$$\tau_1^2 = \frac{6(n-2)}{(n+1)(n+3)} \doteq \frac{6}{n} \ ,$$

$$\tau_2^2 = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)} \doteq \frac{24}{n} \ ,$$

Case (typist/test)	Digraph	n	Mean	Variance	Skewness	Kurtosis
612	s-	27	4.702	0.056	-1.223	0.986
612	th	29	4.930	0.020	1.109	5.379
612	-a	26	4.921	0.027	0.950	2.279
612	-t	27	4.772	0.086	3.085	9.404
622	er	25	5.004	0.023	-0.031	-0.611
622	e-	46	4.928	0.037	0.082	0.523
622	r-	32	4.664	0.092	-0.120	-0.777
622	y-	27	5.072	0.068	1.097	0.739
622	-f	27	4.966	0.057	0.429	-0.759
622	-s	36	5.024	0.053	2.477	8.730
631	d-	25	4.744	0.049	-1.422	2.257
631	en	25	4.924	0.098	1.632	1.732
631	e-	44	4.715	0.055	0.543	0.957
631	he	30	4.803	0.040	0.988	1.122
631	in	28	4.912	0.066	1.672	3.069
631	s-	28	4.791	0.200	-0.543	1.145
631	th	26	4.964	0.015	0.994	1.283
631	-a	27	4.824	0.050	1.080	1.837
631	-t	25	4.800	0.068	0.872	1.363
632	en	25	4.974	0.028	1.269	0.807
632	e-	43	4.826	0.070	0.526	2.283
632	he	29	4.944	0.070	1.009	0.259
632	in	25	4.896	0.030	0.826	0.513
632	s-	28	4.823	0.043	0.115	0.644
632	th	26	5.005	0.047	1.661	3.357

Fig. 5 — Moments of transformed data

$$\hat{\theta} = \hat{\mu}_3/\hat{\mu}_2^{3/2}, \; \hat{\phi} = (\hat{\mu}_4/\hat{\mu}_2^2) - 3$$

$$\hat{\mu}_k = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^k, \; k=2,3,4.$$

A caret over a quantity denotes its value estimated by replacing population quantities by sample quantities. Thus, for n = 25, for example, since the 95 percent fractile for rejection of normality is 2, we should reject at the 5 percent level of significance if

$$|\hat{\theta}| > .88 \text{ , or } |\hat{\phi}| > 1.5 \text{ .}$$

Figure 5 shows that although many sample skewness values exceed .88 and many sample kurtosis values exceed 1.5, the actual sample values are not substantially different from the critical values of .88 and 1.5, respectively. That is to say, while the distributions of the transformed variables are clearly not normal, they appear to be approximately so. The same conclusion holds for all cases, including those not shown. Therefore, we decided to go forward on the assumption that the transformed data were normally distributed.

Figure 6 shows how the distributions compared with one another when all three texts were combined (i.e., digraph times were pooled). The plots for each of the typists were developed for the digraph *th*. There were at least nine replications of each case. While the mean values of the logged digraph times tend to differ from one another, the variances tend to be fairly constant.

CONSISTENCY OF TYPING PATTERNS OVER TIME

The question of whether or not an individual's typing pattern changes over time is a central consideration in determining the feasibility of an authentication method based on keystroke timing.

To investigate typing consistency over time, we studied each digraph separately. For a given typist and a given digraph, there was a set of mutually independent replications available from the August

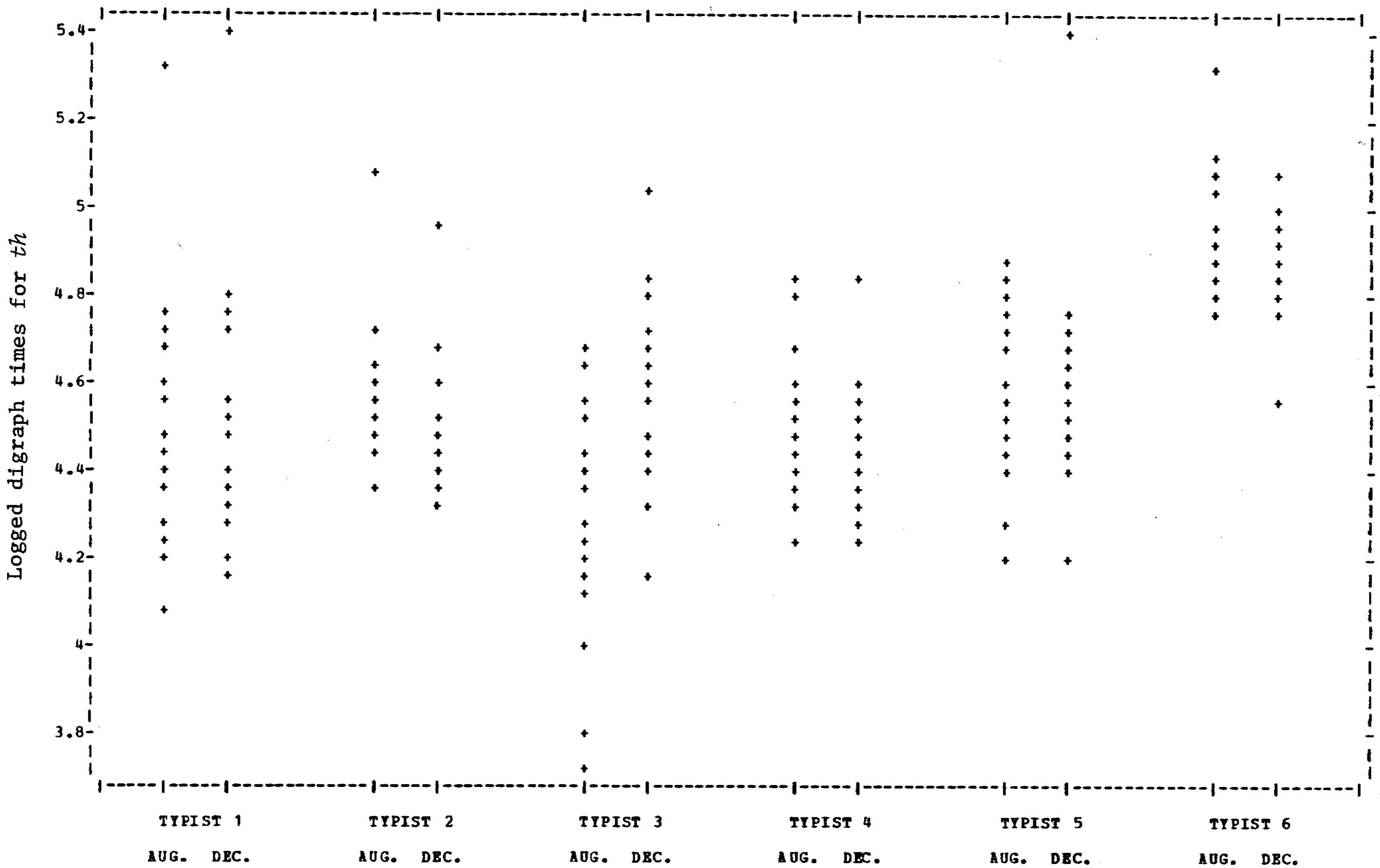

Fig. 6 — Comparison of digraph-time distributions

test, and another set from the December test. (The frequencies of replication of a given digraph for the two tests differ occasionally because of typing errors.)

We took the transformed data as the basic variables, assumed the variances of the transformed digraph times were the same for both sets (we adopted the concept of test robustness explained above), then carried out a classical two-sample Student t-test of the hypothesis that the means of the transformed digraph times were the same. The analysis is given below.

Suppose a given digraph α is replicated by a given typist β, M times in August and N times in December (these are frequencies obtained after removing digraph times exceeding 500 milliseconds). Let the logged values of the digraph times in August be denoted by $X_1,\ldots,X_M$, and the corresponding logged December times be denoted by $Y_1,\ldots,Y_N$. The X_i's and the Y_j's are assumed mutually independent, and independent of one another. Since we purposely logged the data in order to induce normality as an approximating distribution, and because the sample variances in August and December are approximately the same, it is reasonable to assume that

$$X_i \sim N(\theta_1,\sigma^2),\ Y_j \sim N(\theta_2,\sigma^2),$$

$i=1,\ldots M$; $j=1,\ldots,N$. The problem is to test the hypothesis $H:\{\theta_1=\theta_2, \sigma^2>0\}$ versus the alternative hypothesis $A:\{\theta_1\neq\theta_2, \sigma^2>0\}$. If H is true, it implies that β's typing has not changed significantly over the four-month period, insofar as digraph α is concerned. The classical (uniformly most powerful unbiased) test of H versus A is to form the t-statistic

$$t^* = \frac{|\bar{X} - \bar{Y}|}{\sqrt{\eta}\left\{\dfrac{\sum_{i=1}^{M}(X_i-\bar{X})^2 + \sum_{j=1}^{N}(Y_j-\bar{Y})^2}{M+N-2}\right\}^{1/2}}$$

where

$$\eta \equiv \frac{1}{M} + \frac{1}{N}, \quad \bar{X} = \frac{1}{M} \sum_{1}^{M} X_i, \quad \bar{Y} = \frac{1}{N} \sum_{1}^{N} Y_j,$$

and test it for significance, using the fact that under H,

$$t^* \sim t_{M+N-2};$$

that is, under H, t^* follows a Student t-distribution with $M + N - 2$ degrees of freedom. The form of the test is to reject H if the absolute value of t^* exceeds a critical value determined by the significance level of the test.

All digraphs for all typists were compared using this t-test procedure. The three texts were pooled and treated as one text, because there were insufficient frequencies for many digraphs. A sample of our results is shown in Table 3 for those cases in which there were at least ten replications of the digraph in both the August and December tests. For the first entry, with a t-statistic of .351, when H is true we have $P\{t^* > .351\} = .728$. Such a t-statistic is quite likely to have occurred by chance under H, so H cannot be rejected. The digraphs for which the t-statistic is considered significant (for which the p-value is less than .05) are indicated by an asterisk to the right of the entry in the p-value column. The last column in the table shows which hands are used to type each digraph. For example, the first digraph, *ir*, is conventionally typed with a finger of the right hand followed by a finger of the left land; thus, the entry R-L denotes "right" and "left," respectively. We included this information in order to determine if a hand pattern would emerge for those digraph tests that were significant. We could not find any such pattern. Of the 144 cases evaluated for Typist 2 (only 50 of which are shown in Table 3), 7.6 percent were significant. That is, H was rejected about 8 percent of the time, which means that the subject's typing was consistent, from August to December, on 92 percent of the digraphs. The consistencies of the other typists are shown below:

Table 3

RESULTS OF T-TESTS OF TRANSFORMED DIGRAPH-TIME DISTRIBUTIONS
(Typist 2, 144 digraphs, 7.6 percent significant)

Digraph	Replications		Mean Digraph Times			Standard Deviations				t-statistic	p-value	Hand(s) used to type digraph
	Aug	Dec	Aug	Dec	Delta[a]	Aug	Dec	Ratio	Pooled			
ir	17	18	4.687	4.649	0.038	0.342	0.298	1.147	0.321	0.351	0.728	R - L
is	34	22	4.816	4.864	-0.047	0.180	0.229	0.788	0.201	-0.863	0.392	R - L
it	37	22	4.821	4.680	0.141	0.452	0.252	1.799	0.391	1.343	0.184	R - L
iv	21	21	4.725	4.745	-0.021	0.160	0.203	0.792	0.183	-0.367	0.715	R - L
ke	17	14	4.652	4.697	-0.045	0.286	0.289	0.989	0.287	-0.432	0.669	R - L
ki	17	15	5.077	5.089	-0.012	0.161	0.202	0.797	0.182	-0.188	0.852	R - R
k-	13	11	4.807	4.835	-0.028	0.148	0.181	0.822	0.164	-0.412	0.684	R - R
la	30	25	4.683	4.615	0.068	0.152	0.109	1.395	0.135	1.852	0.070	R - L
le	43	31	4.745	4.813	-0.068	0.247	0.218	1.134	0.235	-1.227	0.224	R - L
lf	11	12	4.887	5.136	-0.250	0.413	0.448	0.920	0.432	-1.385	0.181	R - L
ll	22	14	5.023	5.038	-0.015	0.297	0.133	2.240	0.248	-0.176	0.862	R - R
lu	17	14	5.126	5.115	0.011	0.211	0.226	0.933	0.218	0.142	0.888	R - R
ly	15	11	5.261	5.221	0.040	0.214	0.155	1.381	0.191	0.527	0.603	R - R
l-	32	24	4.614	4.478	0.136	0.383	0.394	0.971	0.388	1.303	0.198	R - R
ma	22	11	4.830	4.715	0.114	0.262	0.243	1.076	0.256	1.209	0.236	R - L
me	27	19	4.691	4.568	0.122	0.111	0.153	0.729	0.130	3.149	0.003 *	R - L
mi	14	12	5.125	5.117	0.008	0.171	0.162	1.055	0.167	0.125	0.901	R - R
mp	13	12	4.967	5.115	-0.148	0.196	0.293	0.668	0.247	-1.490	0.150	R - R
ms	14	13	4.883	4.851	0.031	0.197	0.226	0.872	0.211	0.385	0.703	R - L
m-	18	15	4.790	4.861	-0.072	0.255	0.256	0.994	0.255	-0.803	0.428	R - R
nd	41	31	4.564	4.569	-0.004	0.219	0.202	1.084	0.212	-0.086	0.932	R - L
ng	33	17	4.675	4.643	0.033	0.184	0.402	0.457	0.277	0.397	0.693	R - L
no	21	14	5.198	5.144	0.054	0.151	0.131	1.157	0.143	1.086	0.285	R - R
ns	24	13	4.739	4.647	0.092	0.306	0.167	1.832	0.267	1.007	0.321	R - L
nt	28	19	4.749	4.631	0.118	0.401	0.254	1.581	0.349	1.135	0.262	R - L
n-	39	28	4.914	4.878	0.035	0.170	0.174	0.977	0.171	0.834	0.407	R - R
oc	10	10	4.821	4.750	0.072	0.295	0.357	0.826	0.327	0.491	0.629	R - L
of	23	15	4.622	4.608	0.014	0.121	0.196	0.618	0.154	0.271	0.788	R - L
om	24	20	5.088	5.071	0.017	0.228	0.228	0.999	0.228	0.240	0.811	R - R
on	42	32	5.091	5.090	0.001	0.146	0.114	1.280	0.133	0.021	0.983	R - R
or	44	37	4.694	4.723	-0.030	0.199	0.296	0.673	0.248	-0.537	0.593	R - L
ox	10	10	5.067	5.065	0.002	0.147	0.120	1.227	0.134	0.041	0.968	R - L
o-	17	11	4.727	4.731	-0.003	0.258	0.218	1.180	0.243	-0.036	0.971	R - R
pe	21	13	4.746	4.710	0.036	0.425	0.208	2.047	0.359	0.282	0.780	R - L
pl	18	15	5.285	5.271	0.014	0.153	0.065	2.365	0.122	0.341	0.736	R - R
pr	19	16	4.779	4.746	0.033	0.230	0.255	0.900	0.242	0.407	0.687	R - L
ra	45	34	5.106	5.140	-0.034	0.143	0.257	0.558	0.200	-0.756	0.452	L - L
rd	17	14	5.285	5.305	-0.019	0.142	0.182	0.779	0.161	-0.330	0.744	L - L
re	41	33	5.093	5.076	0.017	0.210	0.173	1.215	0.195	0.367	0.715	L - L
ri	16	11	4.543	4.546	-0.003	0.111	0.111	1.000	0.111	-0.068	0.946	L - R
rk	11	11	4.620	4.774	-0.154	0.116	0.177	0.659	0.149	-2.413	0.026 *	L - R
rm	13	14	4.653	4.762	-0.109	0.091	0.319	0.284	0.239	-1.186	0.247	L - R
rn	11	10	4.679	4.809	-0.131	0.095	0.266	0.357	0.196	-1.529	0.143	L - R
ro	23	13	4.707	4.624	0.083	0.201	0.078	2.563	0.168	1.416	0.166	L - R
r-	58	47	4.620	4.647	-0.026	0.149	0.140	1.061	0.145	-0.931	0.354	L - R
se	34	24	5.166	5.114	0.052	0.092	0.091	1.003	0.092	2.140	0.037 *	L - L
sm	20	16	4.938	4.852	0.086	0.291	0.266	1.095	0.281	0.917	0.365	L - R
st	28	20	5.105	5.085	0.020	0.135	0.151	0.893	0.142	0.488	0.628	L - L
sw	16	15	5.403	5.330	0.073	0.165	0.092	1.807	0.135	1.508	0.142	L - L
s-	74	48	4.809	4.724	0.085	0.305	0.191	1.593	0.266	1.727	0.087	L - R

[a]Difference between August and December means.

Typist	Percent of Cases Significant	Consistency (percent)
1	11.7	88.3
2	7.6	92.4
3	50.0*	50.0*
4	4.7	95.3
5	5.5	94.5
6	20.6	79.4

The t-tests of typing consistency over time rest on several model assumptions. The distributional assumption of normality was reasonable in light of the fact that the data were transformed until their distribution was approximately normal. The variances of the transformed digraph-time distributions differed from one digraph to another, but for a given digraph, they varied very little from August to December. For this reason we pooled the data from the two tests insofar as sample-variance computations were concerned. It is well known that t-tests are quite insensitive to small deviations from normality and small excursions of the variance ratio from unity, so we felt quite confident of the results of our test. We therefore concluded that it was reasonable to consider authentication procedures based on keystroke timing, since subjects are likely to be sufficiently consistent in their typing patterns over time for such a procedure to be effective. The results of our authentication analysis are summarized below.

DEVELOPMENT OF AN AUTHENTICATION PROCEDURE

The statistical model developed for authenticating subjects on the basis upon their keystroke timing patterns is presented in the Appendix. In the following, we describe the results of applying this statistical model to the empirical data obtained in our experiment.

Since the digraph frequencies in each of the three separate texts were often very low (too low to permit meaningful statistical inferences),

---

*Typist 3 completed only the August test for Text 1 (see Fig. 1), so these results are based upon only 48 cases. This subject took the tests unenthusiastically and slowed down substantially (but consistently) during her second test.

we pooled the data from all texts typed by a given subject during a given test into a single sample. As indicated in Table 2, this usually meant that three texts were pooled, but in two cases only two texts were pooled, and Typist 3 typed only one text during the August test. The pooling resulted in "reasonably large" digraph frequencies for a number of digraphs for all samples except the August test of Typist 3. Since that sample did not provide sufficiently large numbers of digraph replications for statistical inferences to be made, we excluded it from our subsequent analyses. This left 87 digraphs for which ten or more replications were available in each of the remaining eleven cases (December for Typist 3 and both August and December for the other five typists).

Our first test of authentication used all 87 digraphs. The combination of a given subject and a given time (say, Typist 2, August test) was used to define the "originator." Then all the other tests, in both August and December, were compared with the originator's test. All these others were considered "claimants," including Typist 2 in the December test. Any authentication test results other than those in which Typist 2, December, was authenticated were considered errors (a primary error if originator and claimant were different, but the procedure authenticated; a secondary error if originator and claimant were the same, but the procedure did not authenticate). Since there were eleven cases, there were eleven possible originators and ten possible claimants for each choice of originator. However, the roles of originator and claimant were symmetric in our procedure; that is, when comparing two samples it is irrelevant which of them is labeled originator and which is labeled claimant, as the results will be the same in either case. Thus we had 55 unique authentication tests.

In each authentication test, a vector of transformed means for the 87 digraphs of the originator was compared with a similar vector of transformed means of the same 87 digraphs of the claimant. In each case, we studied both the number of primary and secondary errors made and the p-value corresponding to the strength of the 55 separate tests.

The results for all tests showed no primary errors, although there were two secondary errors (Typists 1 and 6 were both incorrectly denied access to the computer when their August tests were compared against their own December tests). The questions at this point were the following:

1. Is it possible that the secondary errors could be eliminated by eliminating certain digraphs? If so, which digraphs should be eliminated?
2. How small a vector of digraphs can be used for authentication without any primary or secondary errors occurring? Which digraphs are the "key" ones?

We hypothesized a mechanism that might be generating the observed differences in the keystroke-timing patterns, in the hope that such a hypothesis would serve as a guide for eliminating digraphs from the 87-dimensional vector (the alternative would have been to study every possible subset, a very difficult undertaking). We assumed that observed differences occur because of the differences in finger dexterity and muscular coordination between subjects. If this is correct, it is unlikely that using bogus digraphs such as (e, -) would contribute very much to our understanding, since they represent aggregations over the second character in the digraph (see p. 15) which would be likely to mask individual differences. We also reasoned that finger dexterity would most likely be different on different hands of the same subject, so we decided to study authentication patterns using certain finger and hand combinations.

We first eliminated all digraphs that contain a space. When the same 55 authentication tests were carried out on the remaining 60 digraphs, one secondary error occurred: Typist 6 was again denied access to the computer when she should have been authenticated.

Of these 60 digraphs, 11 were made with two right-hand fingers, and 17 were made with two left-hand fingers. The remaining 32 digraphs required a different hand for each of the two characters.

Authentication tests performed with only the 17 left-left (L-L) digraphs produced one primary error and one secondary error. But when we used only the 11 right-right (R-R) digraphs, we found no errors of either kind--*a perfect authentication record.* Therefore, we decided to concentrate on R-R digraphs for authentication.

We next addressed the question of whether or not the factor of size 11 could be reduced. We started by studying the strength with which authentication was carried out in each of the tests when particular digraphs were deleted from the set of 11.* This process ultimately suggested a subset of 5 R-R digraphs with which authentication could be carried out for all 55 tests without primary or secondary errors. This subset comprises a core of four "necessary" digraphs--*in*, *io*, *no*, *on*--plus one other digraph that could be *ul*, *il*, or *ly*; that is, the core plus any one of these three could be used to produce authentication with no errors. These digraphs are all typed using only the second, third, or fourth fingers of the right hand. Note also that each digraph contains at least one vowel (including *y*).

In addition to determining that our authentication procedure will work without error in all cases, it is important to understand the strength of the procedure. That is, when the procedure authenticates in a particular instance, does it do so just barely, or does it do so with very little question? When the procedure says the claimant is an impostor, does it give a resounding rejection or a borderline one?

Our authentication procedure was keyed to operate at a 5 percent level of significance (other significance levels can be selected for a given situation, but we retained this level throughout our preliminary study for convenience and consistency). This means that when claimant is an impostor, the procedure should authenticate with a p-value $\geq .05$. On our tests using the digraphs *in*, *io*, *no*, *on*, and *ul*, the p-values were as shown in Table 4.

In our comparisons, when the p-value should have been large ($\geq .05$), it actually was very large, in all cases except that of Typist 6 versus Typist 6, where the p-value was only .078. In the opposite situation, where we wanted a small p-value ($\leq .05$), it was very small in all cases.

---

*We also studied the rankings for each typist for each digraph and found cases where the ranks for a particular digraph differed strongly across subjects. Such a digraph was considered a candidate for retention, and others were rejected.

Table 4

RESULTS OF AUTHENTICATION PROCEDURE USING
DIGRAPHS *IN*, *IO*, *NO*, *ON*, AND *UL*

Case (Typist/Test)	p-value[a]	Case (Typist/Test)	p-value[b]
1/Aug vs. 1/Dec	.304	1 vs. all others	.017
2/Aug vs. 2/Dec	.321	2 vs. all others	.001
--	--	3 vs. all others	.001
4/Aug vs. 4/Dec	.977	4 vs. all others	.000
5/Aug vs. 5/Dec	.150	5 vs. all others	.017
6/Aug vs. 6/Dec	.078	6 vs. all others	.004

[a]Should be $\geq$ .05.

[b]Should be $\leq$ .05.

The weakest case was that of Typist 6 versus Typist 6, where the p-value was .078. Thus, the procedure worked quite well at the 95 percent confidence level.* In some situations, 90 percent confidence or less would be adequate, while in other, critical situations, 99.999 percent or more might be required. It is likely that situations requiring high levels of confidence would require very sophisticated digraph combinations (or possibly trigraphs or tetragraphs).

We do not yet fully understand why the particular digraphs we studied appear to be the key discriminators among our small sample of subjects, and of course we do not yet know whether these digraphs would serve us as well in a new, different, and larger sample. These preliminary results are sufficiently promising, however, to make us very hopeful for positive results in related research in the future. Because the two left-handed subjects in our sample were "nonfamilial,"

---

*The significance level in Table 4 could have been anywhere from .017 to .078 (instead of .05), and the same results would have been obtained, i.e., there would have been no errors of authentication. This corresponds to a confidence-level variation of from 92.2 to 98.3 percent. This set of five digraphs was our best case in terms of the possible range of error-free confidence-level variation.

that is, left-handedness does not "run" in either of their families, we believe that the organization of the cerebral hemispheres of their brains is similar to that of right-handed people (see Hardyck et al., 1977). That is, the left hemispheres of their brains probably control the typing patterns that are likely to be most subject-specific in the right hand. It is therefore not surprising that all six subjects could be authenticated with R-R digraph combinations only. Future samples should include familial left-handers as well, to determine whether L-L or L-R and R-L digraphs will also be required for authentication.

V. CONCLUSIONS

The results obtained so far in this study have been very gratifying. However, our explorations into this important area of research are very preliminary, and our conclusions are based upon a small and imperfect sample of data. Therefore, we must qualify them in many ways.

Nevertheless, preliminary analysis strongly suggests that there is indeed a typing "signature"; that is, professional typists really do appear to have distinguishable "styles" of typing, as measured by patterns of expected times to type certain digraphs.

The second, and certainly subsidiary, conclusion of this study is that with the statistical authentication procedure we have developed, the five digraphs *in*, *io*, *no*, *on*, and *ul* are sufficient to distinguish right-handed touch typists from one another in a reliable way. This result must of course be validated on new samples of much greater size, and for less expert typists. We are cautiously optimistic that further experimentation will corraborate our preliminary findings.

Appendix

DERIVATION OF THE STATISTICAL MODEL FOR AUTHENTICATION

DEFINITION OF THE PROBLEM

Let $X:(r\times 1)$ denote an $r\times 1$ random vector of observable characteristics corresponding to the keystroke-timing performance of the originator, and $Y:(r\times 1)$ denote an analogous vector for the claimant. Assume X follows the normal probability distribution with mean θ_o and covariance matrix D,

$$\Pi_o = N(\theta_o, D),$$

and that Y follows the distribution

$$\Pi_c = N(\theta_c, D),$$

where D denotes the diagonal matrix

$$D = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_r^2).$$

The raw digraph times corresponding to the originator and claimant are assumed to have been mathematically transformed until they satisfy the above assumptions. Suppose a sample of size M is available from Π_o for the originator, and a sample of size N is available from Π_c for the claimant. That is, we have available independent observation vectors $(x_1, \ldots, x_M)$ and $(y_1, \ldots y_N)$, the two sets are assumed independent, and the x_j's follow Π_o, while the y_k's follow Π_c. The authentication problem is now one of hypothesis testing (in a classical statistical sense) in that we wish to test the hypothesis that $\Pi_o = \Pi_c$, versus the alternative hypothesis that $\Pi_o \neq \Pi_c$.

If the originator and the claimant are statistically the same person, we will conclude that the keystroke-timing characteristics of the claimant are sufficiently similar to those of the originator that we are inclined to conclude with a high degree of confidence that such keystroke patterns were most likely generated by the same individual. If a test of hypotheses suggests that $\Pi_o \neq \Pi_c$, we should conclude that the keystroke-timing patterns of the originator and claimant are sufficiently dissimilar that the subjects are most likely different people. (An alternative explanation for an observed difference is, of course, that the originator and claimant are actually the same person, but for some reason, the keystroke timing "signature" of the subject has undergone a structural change.)

Using conventional statistical notation, we will test the hypothesis

$$H: \theta_o = \theta_c, \ D > 0, \ D \text{ is diagonal}$$

against the hypothesis

$$A: \theta_o \neq \theta_c, \ D > 0, \ D \text{ is diagonal}$$

where the notation D > 0 means that the matrix D is assumed to be any positive definite (symmetric) matrix (and of course, in this instance, it must be diagonal as well).

REDUCTION TO CANONICAL FORM

We now put the problem into canonical form by first going to sufficient statistics. Define the sample means and variances

$$\bar{x} = \frac{1}{M}\sum_{1}^{M} x_i, \ \bar{y} = \frac{1}{N}\sum_{1}^{N} y_j,$$

$$v_k^2 = \sum_{i=1}^{M}(x_{ki} - \bar{x}_k)^2 + \sum_{j=1}^{N}(y_{kj} - \bar{y}_k)^2,$$

where

$$x_i \equiv (x_{ki}), \ y_j \equiv (y_{kj}),$$

$$\bar{x} \equiv (\bar{x}_k), \ \bar{y} \equiv (\bar{y}_k);$$

and let v denote the vector of sample sums of squares:

$$\underset{(r\times 1)}{v} \equiv (v_1^2,\ldots,v_r^2)'.$$

Note that $(\bar{x}, \bar{y}, v)$ is sufficient for (θ_o, θ_c, D).

The distributions of the sufficient statistics are well known. Since

$$\mathcal{L}(\bar{x}) = N(\theta_o, \frac{D}{M}),\ \mathcal{L}(\bar{y}) = N(\theta_c, \frac{D}{N}),\ \mathcal{L}(\bar{x}-\bar{y}) = N(\theta_o - \theta_c, \tau D),$$

where $\tau \equiv (M^{-1} + N^{-1})$, and $(\cdot)$ denotes the probability law of the quantity in parenthesis.

Note that

$$\mathcal{L}\left(\frac{v_i^2}{\sigma_i^2}\right) = \chi_\nu^2$$

where $\nu = M + N - 2$. The problem may now be rewritten in the more compact form:

$$\mathcal{L}(z) = N(\phi, D),\ \mathcal{L}(v_i^2/\sigma_i^2) = \chi_\nu^2\ ,$$

where

$$z \equiv \frac{\bar{x} - \bar{y}}{\sqrt{\tau}}\ ,\ \phi \equiv \frac{\theta_o - \theta_c}{\sqrt{\tau}}\ ,$$

and the problem is to test

$$H\colon \phi = 0,\ D > 0,\ \text{vs. } A\colon \phi \neq 0,\ D > 0.$$

Clearly (z,v) is sufficient for (ϕ,D); also, it is well known that z and v are stochastically independent.

LIKELIHOOD-RATIO TEST

Let Ω denote the parameters in the canonical problem, so that

$$\Omega \equiv (\phi;D) = (\phi;\sigma_1^2,\ldots,\sigma_r^2).$$

The joint density of the sufficient statistics is given by

$$f(z,v|\Omega) = f_1(z|\Omega)\ f_2(v|\Omega),$$

where

$$f_1(z|\Omega) \propto |D|^{-1/2}\ \exp(-1/2)\{(z-\phi)'D^{-1}(z-\phi)\},$$

$$f_2(v|\Omega) = \prod_{j=1}^{r} g(v_j|\sigma_j^2),$$

$$g(v_j|\sigma_j^2) \propto \frac{(v_j^2)^{(\frac{\nu}{2}-1)}}{(\sigma_j^2)^{\frac{\nu}{2}}}\ \exp(-1/2)\left\{\frac{v_j^2}{\sigma_j^2}\right\}.$$

The notation $\propto$ means "is proportional to," the prime denotes a transposed matrix, and $|D|$ denotes the determinant of D. Combining terms shows that we can write

$$f(z,v|\Omega) \propto \prod_{j=1}^{r} h_j(z_j,v_j^2|\phi_j,\sigma_j^2),$$

where

$$z \equiv (z_j),\ v \equiv (v_j^2),\ \phi \equiv (\phi_j),\ \text{and}$$

$$h_j(z_j,v_j^2|\phi_j,\sigma_j^2) = \frac{(v_j^2)^{(\frac{\nu}{2}-1)}}{(\sigma_j^2)^{(\nu+1)/2}}\ \exp\left\{-\left(\frac{1}{2\sigma_j^2}\right)\left[v_j^2 + (z_j-\phi_j)^2\right]\right\}.$$

The likelihood ratio statistic (LRS) for testing H versus A is defined as

$$\lambda = \frac{\max_{H} f(z,v|\Omega)}{\max_{HUA} f(z,v|\Omega)} .$$

It is straightforward to check that in this case, λ is given by

$$\lambda^{-1} = \prod_{j=1}^{r} [1 + z_j^2/v_j^2]^{(\nu+1)/2} .$$

The test is to reject H if λ is too small, i.e., reject H if $\lambda < C^*$, where C^* denotes some constant that must still be determined.

DISTRIBUTION OF LRS UNDER H

Define

$$U \equiv \lambda^{2/(\nu+1)} = [\prod_{1}^{r}(1 + z_j^2/v_j^2)]^{-1} .$$

Then, an equivalent test is to reject H if $U < C$, where C is some unknown constant that must be determined.

Now note from the above distributional statements that under H,

$$\mathcal{L}\left(\frac{z_j^2}{v_j^2}\right) = \mathcal{L}\left(\frac{\chi_1^2}{\chi_\nu^2}\right) ,$$

where χ_1^2 and χ_ν^2 are independent. Thus, from a distributional standpoint, we may write

$$U = \prod_{j=1}^{r} Z_j ,$$

where

$$\mathcal{L}(Z_j) \equiv \mathcal{L}\left(\frac{\chi_\nu^2}{\chi_\nu^2 + \chi_1^2}\right).$$

Next note that $\mathcal{L}(Z_j) = \beta(\frac{\nu}{2}, \frac{1}{2})$, where $\beta(a,b)$ denotes a beta distribution with density

$$p(x|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1, \ a > 0, \ b > 0,$$

and is 0 otherwise; and $B(a,b)$ denotes a beta function. Thus, the test statistic U is distributed under H as the product of independent beta variates with identical degrees of freedom.

APPROXIMATE DISTRIBUTION OF LRS UNDER H

The exact distribution of a product of independent beta variates is very complicated. We therefore propose below an approximation which is adequate for our purposes. This approximation involves replacing the product of independent beta variates by a single beta variate that has the same first two moments (see Tukey and Wilks, 1946).

Accordingly, assume

$$\mathcal{L}(U) \cong \beta(\gamma,\delta),$$

where (γ,δ) are degrees-of-freedom parameters that will be determined in terms of the known constants (ν,r). Note that since

$$U = \prod_{j=1}^{r} Z_j,$$

and because it is well known that

$$E(U) = \frac{\gamma}{\gamma+\delta} \quad ,$$

$$\frac{\gamma}{\gamma+\delta} = E(\prod_1^r Z_j) = \prod_1^r EZ_j = \prod_1^r \left(\frac{\nu/2}{\nu/2 + 1/2}\right)$$

or

$$\left(\frac{\gamma}{\gamma+\delta}\right) = \left(\frac{\nu}{\nu+1}\right)^r . \tag{1}$$

Similarly, since

$$E(U^2) = E(\prod_1^r Z_j^2) = \prod_1^r E(Z_j^2),$$

and it is well known that

$$E(Z_j^2) = \frac{\Gamma(\frac{\nu+4}{2})\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\quad(\frac{\nu+5}{2})} = \frac{\nu(\nu+2)}{(\nu+1)(\nu+3)},$$

$$\frac{\gamma(\gamma+1)}{(\gamma+\delta+1)(\gamma+\delta)} = \left[\frac{\nu(\nu+2)}{(\nu+1)(\nu+3)}\right]^r . \tag{2}$$

Equations (1) and (2) must now be solved simultaneously for (γ,δ), for fixed (ν,r). Define the constants

$$w_1 \equiv (\frac{\nu}{\nu+1})^r,\ w_2 \equiv \left[\frac{\nu(\nu+2)}{(\nu+1)(\nu+3)}\right]^r .$$

It is straightforward, though tedious, to show that

$$\gamma = \frac{w_1(w_1-w_2)}{(w_2 - w_1^2)},\ \delta = \frac{(1-w_1)(w_1-w_2)}{(w_2 - w_1^2)} . \tag{3}$$

It is also straightforward to check that $\gamma > 0$, $\delta > 0$.

AUTHENTICATION TEST

The test for authentication is the test of hypothesis H versus A. That is, if we cannot reject H, we conclude that the claimant should be authenticated; otherwise, we conclude that the claimant is an impostor. The test of H versus A developed above is to reject H if $U < C$, where C was not yet determined. Now we know that under H, it is approximately true (for all sample sizes) that $\mathcal{L}(U) = \beta(\gamma,\delta)$. Therefore,

$$P\{U < C|H\} = F(C),$$

where $F(C)$ denotes the cumulative distribution function of a beta variate with (γ,δ) degrees of freedom; i.e.,

$$F(C) = \frac{1}{B(\gamma,\delta)} \int_0^C x^{\gamma-1}(1-x)^{\delta-1}dx \; .$$

$F(C)$ is also known as an incomplete beta function. Let $\alpha \equiv F(C)$ denote the level of significance of the test of the hypothesis. If α is prespecified according to the level of risk the decisionmaker is willing to take (the size of α will vary according to the context of the problem), since $F(C)$ is a monotone function of its argument, C will be uniquely determined.

The test for authentication now becomes: Do not authenticate if $U < C$, and authenticate if $U \geq C$, where

$$U = \frac{1}{\prod_{k=1}^{r} (1 + t_k^2)} \; ,$$

and

$$t_k^2 = \frac{(\bar{x}_k - \bar{y}_k)^2}{v_k^2} \cdot \left(\frac{MN}{M+N}\right),$$

$$v_k^2 = \sum_{i=1}^{M}(x_{ki} - \bar{x}_k)^2 + \sum_{j=1}^{N}(y_{kj} - \bar{y}_k)^2 \; .$$

BIBLIOGRAPHY

Coover, J. E., "A Method of Teaching Typewriting Based on a Psychological Analysis of Expert Typing," *National Educational Association, Addresses and Proceedings*, Vol. 61, 1923, pp. 561-567.

Cramér, H., *Mathematical Methods of Statistics*, Princeton University Press, Princeton, New Jersey, 1946, pp. 357, 366, 386.

Dvorak, A., N. L. Merrick, W. L. Dealey, and G. C. Ford, *Typewriting Behavior*, American Books, New York, 1936.

Harding, D. W., "Rhythmization and Speed of Work," *British Journal of Psychology*, Vol. 23, 1933, pp. 262-278.

Hardyck, C., and Lewis F. Petrinovich, "Left-Handedness," *Psychological Bulletin*, Vol. 84, No. 3, 1977, pp. 385-404.

Lahy, J. M., *Motion Study in Typewriting*, World Peace Foundation, Boston, 1924.

Neal, Alan S., "Time Intervals Between Keystrokes, Records, and Fields in Data Entry with Skilled Operators," *Human Factors*, Vol. 19, No. 2, 1977, pp. 163-170.

Ostry, David Joseph, "The Organization of Typewriting Performance," Ph.D. thesis, Department of Psychology, University of Toronto, 1977.

Rochester, N., F. T. Bequaert, and E. M. Sharp, "The Chord Keyboard," *Computer*, December 1967, pp. 57-63.

Tukey, J., and S. Wilks, "Approximation of the Distribution of the Product of Beta Variables by a Single Beta Variable," *Annals of Mathematical Statistics*, Vol. 17, 1946, pp. 318-324.