社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2007－ＣＳＥＣ－38（４）
2007／7／19

キーストロークの統計情報を利用した個人認証手法の提案

片岡 祥啓† 宮本 貴朗†† 青木 茂樹†† 泉 正夫† 福永 邦雄†

† 大阪府立大学 大学院工学研究科
†† 大阪府立大学 総合教育研究機構
〒 599-8531 大阪府堺市中区学園町 1-1
E-mail: †kataoka@com.cs.osakafu-u.ac.jp, ††{aki, aoki}@las.osakafu-u.ac.jp,
†{izumi, fukunaga}@cs.osakafu-u.ac.jp

あらまし これまでに提案されているキーストロークを利用した個人認証の研究では，キーが押されていた時間を比較したり，パスワードを打つリズムを意図的に作り出すことによってユーザ認証を行っていた．しかし，これらの手法はパスワードなどの定型文には有効であるが，自由な文章（非定型文）を用いた場合は，定型文に比べ大幅に減少するため識別率が大きく低下するという問題があった．本稿では，非定型文を用いた場合にもユーザの特徴を的確に捉え，個人認証を行うことができる手法を提案する．まず，キーストロークの特徴として，連続する二文字（二連字）を押した時の時刻と離した時の時刻から計算できる各種の時間を特徴として抽出する．次に，二連字の種類ごとにその特徴の平均値で昇順に並べる．そして，その並び方と予め作成しておいたプロファイルの並び方を Kendall の順位相関係数で比較する．さらに，二連字が押されていた時間をプロファイルデータに登録されている時間と比較することにより個人を識別する．

キーワード バイオメトリクス，キーストローク，個人認証

User Authentication Method Based on Statistical Analysis of Keystrokes

Yoshihiro KATAOKA†, Takao MIYAMOTO††, Shigeki AOKI††, Masao IZUMI†, and Kunio FUKUNAGA†

† Graduate School of Engineering, Osaka Prefecture University
†† Faculty of Liberal Arts and Science, Osaka Prefecture University
Gakuen-cho 1-1, Naka-ku, Sakai, Osaka, 599-8531 Japan
E-mail: †kataoka@com.cs.osakafu-u.ac.jp, ††{aki, aoki}@las.osakafu-u.ac.jp,
†{izumi, fukunaga}@cs.osakafu-u.ac.jp

Abstract In this paper, we propose a new user authentication method based on statistical analysis of keystrokes. First of all, we measure the time a key pressed and the time a key released, and compute the feature of keystrokes using digraph (typing keys of two successive letters). Next, we sort the typing samples of the digraphs by their duration, and compare the sorted samples with the profile data collected in advance by using kendall tau rank correlation coefficient. We have performed experiments by this method and have attained good results.

Key words Biometrics, Keystrokes, User authentication

1. まえがき

近年，国内の急速なブロードバンド化と Web サービスの充実により，コンピュータの企業や学校，家庭への普及が急速に進んでいる．それに伴い，コンピュータは複雑な計算だけでなく，情報の管理や文章の作成などに多用されるようになり，社会において必要不可欠なものとなっている．しかしながら，コンピュータが様々な情報を持つようになるにしたがって，機密情報を狙ったコンピュータへの不正アクセスが急増し，大きな社会問題となっている．

このような不正アクセスからコンピュータの持つ機密情報を守る手段として，ユーザ認証システムがある．現在，コンピュータのユーザ認証はIDとパスワードの組合せによるパスワード認証が一般的であるが，パスワードが外部に漏洩すると不正アクセスが容易に行なわれてしまうという問題点がある．

そこで，近年注目を浴びているのがバイオメトリクス[1]を利用したユーザ認証システムである．バイオメトリクスを利用したユーザ認証は，主に2つのカテゴリに分けられる．

一つは指紋や網膜などの生体的特徴を利用した認証方法である．この認証方法には，認証精度が高い，パスワードを記憶する必要がないなどの利点があるが，生体的特徴を提供することに抵抗感を感じるユーザが存在したり，外部に指紋[2],[3]や網膜[4]などのユーザ認証に利用する生体情報が漏洩した場合[5]には不正使用の防止が困難であるなどの問題点がある．

もう一つはキーボード操作[6]〜[9]やマウス操作[10]など，他人が模倣することが困難な行動的特徴を利用した認証方法である．この認証方法には，指紋読み取り装置などの特別な装置が必要でなく，容易に認証を行えるなどの利点があるが，ユーザの特徴を的確に捉えることが困難であるという問題点がある．さらに，パスワードなどの定型文を用いてユーザ認証を行った場合よりも，非定型文を用いてユーザの認証を行った場合には認証精度が低下するという問題点もある．

これらのユーザ認証システムは，ログイン時に一度だけ認証を行なうものが多く，一度コンピュータへの不正侵入を許してしまうと不正使用者を発見することが困難であるという問題点もある．そこで本稿では，ログイン後のユーザのキーボード操作を監視して継続的にユーザ認証するために不可欠である，非定型文を利用した場合にも高精度な個人認証が行なえる手法を提案する．

以降，2.節では従来のユーザ認証システムとその問題点について述べ，3.節ではキーストロークの特徴を用いた提案手法について述べる．4.節では実験と考察を行ない，5.節でまとめとする．

2. 従来のユーザ認証システム

2.1 パスワードによるユーザ認証

現在，コンピュータへのログイン時やWebサービスの利用時のユーザ認証には，パスワード認証が一般的に使用されている．このパスワード認証は，ユーザIDとパスワードが一致すれば正規ユーザとして識別され，一致しなければ不正ユーザとして識別される．そのため，導入が非常に容易で，正しいユーザIDとパスワードを用いれば認証に失敗することがないという利点がある．

しかし現在では，コンピュータへのログイン時だけに限らず，数多くのWebサービスの利用時にパスワード認証が必要となり，その全てのパスワード認証において異なるパスワードを設定し，記憶するのは難しい．また，キーボードの操作履歴を記憶するスパイウェアも存在することから，パスワードが外部に漏洩する危険性も高く，パスワードが漏洩した場合には，不正ユーザの侵入を防止することが難しいという問題点もある．

2.2 バイオメトリクスによるユーザ認証

2.2.1 生体的特徴を利用した認証

生体的特徴を利用した認証方法の代表的な例としては，指紋や静脈，網膜などを利用する方法などがあげられる[2]〜[4],[11]．近年，特に利用されることが多い指紋認証を例にとった場合，指紋はユーザ間の差が明確なためユーザの識別が容易で，時間的変化にも強いという利点があるが，指紋を読みとるための装置が必要であったり，グミ指などの偽物の指による指紋偽造が可能であるため，一度漏洩すると不正ユーザの侵入を防止することができないという問題点がある[5]．また，怪我などによって生体的特徴が損なわれた場合には，認証を行なうことができないという問題点もある．網膜や静脈認証においても指紋認証と同様の問題がある．

2.2.2 行動的特徴を利用した認証

行動的特徴を利用した認証方法の代表的な例としては，キーストロークやマウス操作[10]など他人が模倣することが困難な個人の癖や特徴を用いたものがあげられる．キーストロークを用いた従来のユーザ認証システム[6]〜[9]は，指紋や網膜などを利用した認証システムとは異なり，特別な装置を必要としないという利点がある．しかし，キーストロークなどの行動的特徴は，そのときのユーザの心理状態や体調により変化しやすいため，ユーザの癖や特徴を的確に捉えることが難しく，識別が必要な人数が増加するにつれ，正確な認証を行うことが難しい．

これらの問題に対し，例えばキータイプのリズムを特徴として利用した研究[6]では，ユーザIDやパスワードの入力の際に，ユーザが意図的に一定のリズムをつけることによりユーザ認証の精度を向上させている．しかし，この手法ではパスワードなどの短い定型文を利用する場合には有効であるが，今回対象とする非定型文を利用する場合には認証精度が低下するという問題があった．

また，これまでの研究では定型文を使用した場合の検証例が多く，非定型文を使用した場合には，本人認証率や他人拒否率が低下することが多い．この要因として，非定型文では定型文よりも比較できる特徴が少なく，ユーザの特徴抽出が難しいためであると考えられる．

2.3 従来のユーザ認証システムの問題点

従来のユーザ認証システムの問題点として，使用者が途中で席を外したときにコンピュータの不正使用が行なわれたり，何らかの方法でユーザ認証を突破してコンピュータに不正アクセスが行われた場合，その不正使用者を発見することは難しい．この問題点を解決するために，認証をログイン時だけでなく，ログイン後にも継続的に行なう必要があると考えられる．しかし，一定時間が経過するごとにパスワードを入力し，認証を行なうことは現実的でない．

そこで本稿では，ログイン後のユーザのキーボード操作を監視して継続的にユーザ認証するために不可欠である，非定型文を利用した場合にも高精度な個人認証が行なえる手法を提案する．まず，キーストロークの特徴として，連続する二文字（二連字）を押した時の時刻と離した時の時刻から計算できる時間を基にキーストローク特徴を抽出する．次に，二連字の種類ご

とにその特徴の平均値で昇順に並べる．そして，その並び方と予め作成しておいたプロファイルの並び方を Kendall の順位相関係数で比較する．さらに，二連字が押されていた時間をプロファイルデータに登録されている時間と比較することにより個人を識別する．

3. ユーザ認証手法の提案

3.1 キーストローク特徴

本稿では，キーを押したときの時刻（key down）と離したときの時刻（key up）から計算できる時間をキーストローク特徴として抽出する．個人によってキーストローク特徴は異なると考えられるため，本稿ではこの特徴に着目することで個人の識別を行う．

例えば「X」という文字を異なる 2 人のユーザ A と B がキー入力した場合，どちらのユーザによって入力されたのかを入力された文字だけでは識別することはできないが，ユーザ A がキーを押している時間が長く，ユーザ B がキーを押している時間が短いという特徴があることを知っていれば，キーストロークから「X」の文字がどちらのユーザによって入力されたのかを推定することができる．

しかし，単一の文字が入力されていた時間だけでは，ユーザ特徴を的確に捉えることは困難である．図 1 は異なるユーザ A とユーザ B が約 300 文字の文章を入力したとき，それぞれのキーが押されていた平均時間を表している．x 軸は入力された文字を示し，y 軸はキーが押されてから離されるまでに掛かった平均時間 (ms) を示す．ただし，入力していない文字については表示していない．また，ユーザ A が文章 1 と文章 2 を入力した際に得られたデータをそれぞれ A1，A2，ユーザ B が文章 1 を入力した際に得られたデータを B1 とする．図から分かるように，同じユーザ A のデータでも文章が異なれば特徴を的確に捉えることは難しく，またユーザ A とユーザ B を明確に分類できる特徴を発見することも難しい．このことから，一文字ごとの比較ではユーザの特徴抽出が困難なことが分かる．そこで本稿では，単一の文字ではなく連続する二文字（二連字）を使用してキーストロークの特徴抽出を行う．

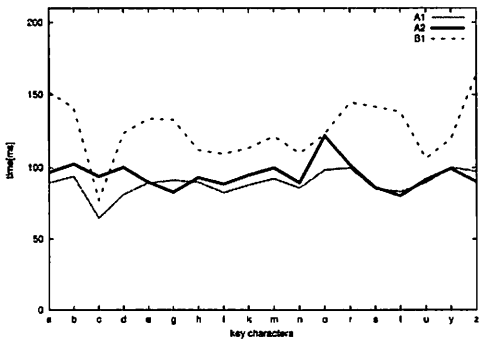

図 1 タイプ時間の比較（一文字）

二連字を用いたキーストローク特徴は，従来手法 [8] に従って定義する．例として ab という二連字が n 回入力された場合，その k 番目のキーストローク特徴を D_{ab}^{k} として，キーストローク特徴 D_{ab} は以下のように定義する．

$$D_{ab} = \frac{1}{n}\left(\sum_{k=1}^{n} D_{ab}^{k}\right) \tag{1}$$

$$D_{ab}^{k} = DD_{ab}^{k} + UU_{ab}^{k} + UD_{ab}^{k} + DU_{ab}^{k} \tag{2}$$

- DD_{ab} (Down-Down Time)
  キー a を押してからキー b を押すまでの時間間隔
- UU_{ab} (UP-UP Time)
  キー a を離してからキー b を離すまでの時間間隔
- UD_{ab} (UP-Down Time)
  キー a を離してからキー b を押すまでの時間間隔
- DU_{ab} (Down-UP Time)
  キー a を押してからキー b を離すまでの時間間隔

ユーザを識別するための統計データ（以下，プロファイルデータという）として，ユーザごとに表 1 のように二連字の種類ごとの平均時間 D_i を保持する．

表 1 プロファイルデータの一例

文字	D_i
ab	D_{ab}
bc	D_{bc}
cd	D_{cd}
⋮	⋮

3.2 認証手法の概要

ここでは，二連字の入力に要した時間 D_i に着目して認証を行う 2 つの提案手法の概要について説明する．まず一つ目の手法は，プロファイルデータと識別を行うデータ（以下，入力データという）の二連字をそれぞれ D_i の値で昇順に並べ，二連字の並び方を評価することによりユーザの識別を行う手法（以下，ord 手法という）である．二つ目の手法は，プロファイルデータと入力データの D_i の値の差を評価することによりユーザの識別を行う手法（以下，abs 手法という）である．本稿では，この 2 つの手法を組み合わせることによりユーザの認証を行う．

2 つの手法の組み合わせ方を図 2 に示す．この認証手法では，始めに認証を行いたい入力データを ord 手法により識別する．次に，ord 手法で不正ユーザのデータと識別された入力データの二連字の並び方を変更し，再び ord 手法を適用して正規ユーザのデータかどうかを識別する．最後に，上記の手法で正規ユーザのデータとして識別された入力データに対し，abs 手法を適用して入力データの識別を行う．

3.3 順位相関性を利用した手法

多くの従来手法では，プロファイルデータと入力データの二連字の D_i の値の差だけを評価することによりユーザの識別を行っている．しかし，この評価だけでは識別するユーザ数が多くなった場合，ユーザの特徴をうまく捉えることが難しく，正確な認証を行えない．そこで，ユーザ数が増加しても的確に

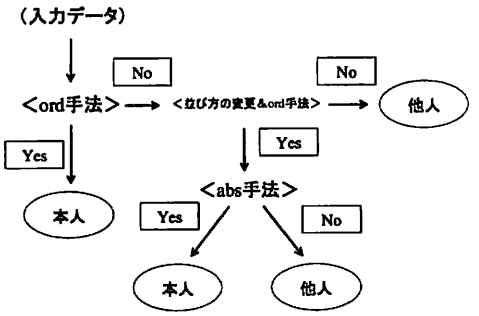

図 2　認証手法の概要

表 2　相関係数の計算例 1(単位：ms)

データ名	プロファイルデータ a		入力データ w	
順位	二連字	D_i	二連字	D_i
1	*aa*	482	*ab*	500
2	*ab*	499	*aa*	511
3	*ac*	501	*bc*	568
4	*ba*	503	*ca*	572
5	*bc*	562	*ac*	579
6	*bb*	563	*cc*	623
7	*ca*	598	*ba*	664
8	*cc*	605	*cb*	670
9	*cb*	672	*ad*	671
10	*ad*	693	*bb*	672

ユーザの特徴を捉えることができるよう識別指標を以下のように定義する．

まず，ユーザ a のプロファイルデータと入力データ w において共通の二連字を取り出し，二連字をそれぞれのデータごとに D_i の値で昇順に並べる．そしてその列び方の類似性を相関係数により評価し，入力データとプロファイルデータの人物が同一人物であるかを判断する．ここでは相関係数の計算式に，Kendall 法[12] を用いる．共通の二連字数を n，入力データ w の i 番目の順位を w_i とすると P_i，Q_i はそれぞれ，$w_i < w_j (j = i+1, i+2, ..., n)$ の個数，$w_i > w_j (j = i+1, i+2, ..., n)$ の個数となる．つまり，$\sum P_i$ は 2 変数の順位の方向が一致する回数，$\sum Q_i$ は 2 変数の順位の方向が逆方向に一致する回数なので，$\sum_{i=1}^{n} P_i - \sum_{i=1}^{n} Q_i$ は順序の一貫性の指標となる．よって，順位相関係数 τ の算出式は以下の通りとなる．

$$\tau = \frac{\sum_{i=1}^{n} P_i - \sum_{i=1}^{n} Q_i}{\frac{n(n-1)}{2}} \qquad (3)$$

例として，表 2 に示すプロファイルデータ a と入力データ w に対して順位相関係数 τ を求めてみる．入力データの二連字をプロファイルデータの二連字の並び方に合わせて並び変え，P_i，Q_i を計算すると表 3 のようになる．例えば，二連字 aa の欄について考える．$n = 10$，$w_1 = 2$ なので $w_1 < w_j (j = 2, 3, ..., n)$ となる個数は 8 となり $P_1 = 8$ となる．また，$w_1 > w_j (j = 2, 3, ..., n)$ となる個数は 1 なので $Q_1 = 1$ となる．これを ab，ac と繰り返していくと $\sum P_i$ と $\sum Q_i$ はそれぞれ 35 と 10 となり，これを式 (3) に代入すると $\tau = (35 - 10)/(10 \cdot 9/2) = 0.555555$ となる．

次に相関係数を基に本人かどうかを判断するための閾値について考える．図 3 は，ユーザ A の入力データ $w^{(1)}$ を識別する場合の閾値の決定方法を示す．まず，入力データ $w^{(1)}$ と残りの入力データ $w^{(2)} \sim w^{(5)}$ をプロファイルデータとして相関係数 τ を算出し，残りの入力データ $w^{(2)} \sim w^{(5)}$ についても同様に相関係数 τ を算出する．そして，相関係数 $\tau^{(2)} \sim \tau^{(5)}$ の平均 μ と標準偏差 σ を使い閾値を $\mu + 2\sigma$ とした．また Kendall の順位相関係数は，0.5 以上で正の強い相関，0.5 以下で正の弱い相関があるとされることから，相関係数が 0.5 以下となる場合はその入力データは不正なユーザのデータと判断することとした．

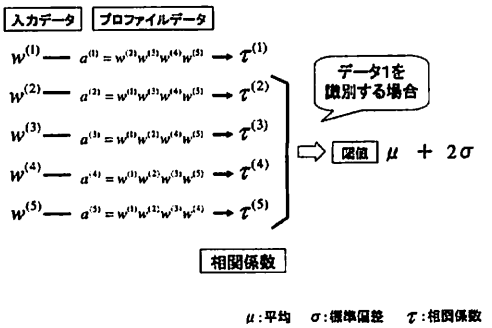

図 3　閾値の決定方法

3.4　二連字の並べ方

3.3 節で述べた手法では，プロファイルデータと入力データの二連字の D_i の値を直接評価せず，二連字の列び方という特徴だけを利用してユーザの識別を行う手法について述べた．しかし，二連字の D_i の値がほとんど変わらない場合でも，順位が大きく変わってしまうという問題点があった．そこで二連字の並べ方を以下のアルゴリズムにより変更することで上記の問題点を解決する．

（1） D_i の値の昇順に並んでいる入力データの二連字に対し，階層的クラスタリングである最短距離法を用いてクラスタリングを行う

（2） クラス同士の最短距離が閾値以上になるまでクラスタリングを続ける

（3） プロファイルデータと比較し，最も相関が強くなるようにクラス内で二連字の順番を入れ換える

（4） ord 手法と同様に，順位相関係数を計算する

（5） 閾値も ord 手法と同じものを利用する

この並び方の変更手法について図 4 を使い具体的に説明する．図 4 は，表 2 のプロファイルデータと入力データの上位 5 つを並べた図である．まず入力データの文字データが ab-aa-bc-ca-ac と並んでいるのに対して，最短距離法を用い，ab-aa

表 3　相関係数の計算例 2

i	1	2	3	4	5	6	7	8	9	10	
二連字	*aa*	*ab*	*ac*	*ba*	*bc*	*bb*	*ca*	*cc*	*cb*	*ad*	
a の順位	1	2	3	4	5	6	7	8	9	10	
w の順位	2	1	5	7	3	8	4	9	10	6	
P_i	8	8	5	3	5	2	3	1	0		$\sum P_i = 35$
Q_i	1	0	2	3	0	2	0	1	1		$\sum Q_i = 10$

と bc-ca-ac との 2 つのクラスに分類する．次に，クラス内で ab-aa を aa-ab，bc-ca-ac を ac-bc-ca とプロファイルデータに対して最も相関係数が大きくなるように並び変えて，再び相関係数を計算する．

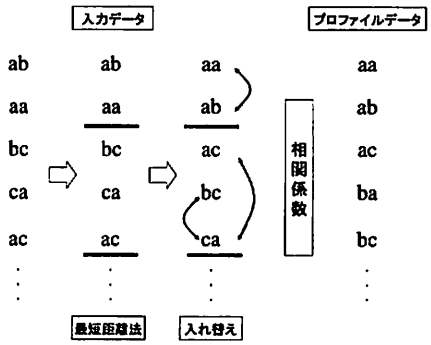

図 4　並べ方の変更例

3.5　打鍵間時間を利用した手法

3.4 節で述べた手法により，順位相関係数の問題点を補うことはできたが，並べ方を変更することにより，これまで拒否できていた他人のデータも誤って認証してしまうという問題がある．

そこで，プロファイルデータと入力データの二連字の D_i の値の差を直接評価することにより最終的なユーザの識別を行うこととする．多くの従来手法では，標準偏差を利用してデータの分散度合に応じて変化するマハラノビス距離を識別に使用しているが，少ないデータ数から正確な標準偏差を求めるのは困難である．換言すると，今回の手法では非定型文を対象としているため，共通の二連字数が少なく，有効なマハラノビス距離を得ることができないと考えられる．

そこで本稿では，従来手法で使われているような標準偏差を使って算出するマハラノビス距離ではなく，標準偏差を利用しないユークリッド距離を使用して D_i の値を評価する．

つまり，プロファイルデータ a と入力データ w に含まれる共通の二連字数を n，プロファイルデータ a の二連字 i における D_i の値を $D_i(a)$，入力データ w の二連字 i における D_i の値を $D_i(w)$ とすると a と w の距離 $D_{feat}(a,w)$ は

$$D_{feat}(a,w) = \frac{1}{n}\sum_{i=aa}^{zz} |D_i(a) - D_i(w)| \qquad (4)$$

となる．

本人かどうかを判断するための閾値については 3.3 節同様，図 3 のように残り 4 つのデータの相関係数の平均 μ と標準偏差 σ を使い閾値を $\mu+\sigma$ とした．

4.　実験と考察

4.1　実験環境

実験では定型文と非定型文の 2 種類に対して，それぞれ実験を行なった．実験に使用する定型文のデータとして，ローマ字入力で約 300 文字（日本語文章では約 200 文字）の日本語文章を 10 人の被験者に 5 回ずつ入力させ，全部で 50 個のタイピングデータを得た．また非定型文のデータとして，ローマ字入力で約 300 文字（日本語文章では約 200 文字）の日本語文章を 13 人の被験者に 5 回ずつ入力させ，全部で 65 個のタイピングデータを得た．

定型文のデータの収集はテキスト文章を画面に表示し，それを見ながら入力させた．また，データの収集は数週間かけて行ない，収集期間に個人差が出ないよう考慮した．このデータには，タイプミスに気づき修正を加えたものとタイプミスに気づかず修正を加えていないものがあった．そのうち修正されなかったタイプミスの部分はタイプミスされた二連字のまま識別を行なった．また，UD の時間間隔が 300ms 以上のデータは修正が加えられたデータと判断し，それ以外にも読点，句読点，スペース，バックスペース，リターンキーを含むキーストロークデータはデータから除外した．

次に実験方法について説明する．まず，定型文を用いた実験では，5 つのタイピングデータのうち一つを認証に使用する入力データ，残りの 4 つをプロファイルデータとして，全てのユーザに対し認証を試みる．以上の処理を全ての入力データに対し行ない，合計 50 回の認証実験を行なった．非定型文を用いた実験でも，定型文を用いた実験同様，5 つのタイピングデータのうち一つを認証に使用する入力データ，残りの 4 つをプロファイルデータとして，全てのユーザに対し認証を試みる．以上の処理を全ての入力データに対し行ない，合計 65 回の認証実験を行なった．

また，実験に使用したデータおよび提案手法で用いたパラメータの値を表 4 に示す．

4.2　定型文に対する提案手法の実験結果

定型文に対して入力データの使用文字数を変化させて提案手法を適用した結果を以下に示す．今回の実験では，残った 4

表 4　パラメータ一覧

パラメータ名	値
実験に使用したデータ	UD が 300ms 以下
閾値 (ord 手法)	$\mu+2\sigma$ or 0.50
閾値 (並び方を変更後)	上記に同じ
最小クラスタ間距離	クラス間が 30ms 以上
閾値 (abs 手法)	$\mu+\sigma$

つの入力データをプロファイルデータとした．実験結果は，正しい利用者本人を拒否する確率として本人拒否率 (FRR:False Reject Rate) と他人を本人と認識してしまう確率として他人受け入れ率 (FAR:False Accept Rate) で表した．

表 5 は ord 手法だけで識別した結果である．入力文字数は実験に使用した入力データの文字数を表す．例えば入力文字数が 50 文字の場合，入力データのうち，最初から 50 番目の文字までを実験に使用した．また，本人成功数，本人失敗数はそれぞれ本人の認証が成功した数と失敗した数を表し，他人受け入れ数，他人拒否数はそれぞれ他人のなりすましが成功した数と失敗した数を表す．次に，二連字の並べ方を変更した後に ord 手法を適用して識別した結果を表 6 に，さらにこの結果に abs 手法を適用して識別した結果を表 7 に示す．

表 5　ord 手法の適用結果 (定型文)

入力文字数	300	200	100	50
本人成功数	49	45	46	40
本人失敗数	1	5	4	10
FRR	2.00%	10.00%	8.00%	20.00%
他人受け入れ数	2	7	12	40
他人拒否数	448	443	438	410
FAR	0.44%	1.56%	2.67%	8.69%

表 6　並べ方を変更後の ord 手法の適用結果 (定型文)

入力文字数	300	200	100	50
本人成功数	50	50	50	48
本人失敗数	0	0	0	2
FRR	0.00%	0.00%	0.00%	4.00%
他人受け入れ数	18	62	114	136
他人拒否数	432	388	336	314
FAR	4.00%	13.78%	25.33%	30.22%

表 7　ユーザ認証実験結果 (定型文)

入力文字数	300	200	100	50
本人成功数	49	47	47	44
本人失敗数	1	3	3	6
FRR	2.00%	6.00%	6.00%	12.00%
他人受け入れ数	2	7	13	40
他人拒否数	448	443	437	410
FAR	0.44%	1.56%	2.89%	8.89%

4.3　非定型文に対する提案手法の実験結果

非定型文に対して行った実験結果を表 8〜10 に示す．表 8 は ord 手法だけで識別した結果である．次に，二連字の並べ方を変更した後に ord 手法を適用して識別した結果を表 9 に，さらにこの結果に abs 手法を適用して識別した結果を表 10 に示す．

また，ユーザ認証実験の結果として図 5 に定型文の FRR, FAR，非定型文の FRR，FAR を示す．

表 8　ord 手法の適用結果 (非定型文)

入力文字数	300	200	100	50
本人成功数	55	54	52	52
本人失敗数	10	11	13	13
FRR	15.38%	16.92%	20.00%	20.00%
他人受け入れ数	38	84	84	116
他人拒否数	742	696	696	564
FAR	4.87%	10.77%	10.77%	14.87%

表 9　並べ方を変更後の ord 手法の適用結果 (非定型文)

入力文字数	300	200	100	50
本人成功数	65	64	63	64
本人失敗数	0	1	2	1
FRR	0.00%	1.54%	3.08%	1.54%
他人受け入れ数	245	267	337	403
他人拒否数	535	513	443	377
FAR	31.41%	34.23%	43.21%	51.67%

表 10　ユーザ認証実験結果 (非定型文)

入力文字数	300	200	100	50
本人成功数	62	59	55	56
本人失敗数	3	6	10	9
FRR	4.62%	9.23%	15.38%	13.85%
他人受け入れ数	44	86	94	124
他人拒否数	736	694	686	656
FAR	5.64%	11.03%	12.05%	15.90%

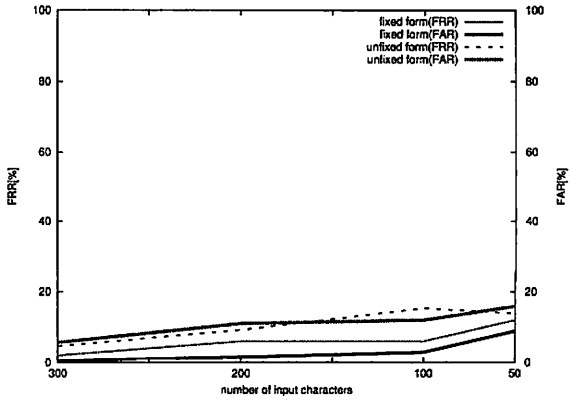

図 5　ユーザ認証実験結果

4.4 考　　察

定型文を用いたユーザ認証では，表7より，入力文字数が50文字のときにはFRRが12.00%，FARが8.89%と認証精度に問題があったが，入力文字数が増加するにつれ認証精度も向上し，入力文字数が300文字のときにはFRRが2.00%，FARが0.44%と良好な結果が得られた．

非定型文を用いたユーザ認証でも，表10より，入力文字数が50文字のときにはFRRが13.85%，FARが15.90%と認証精度に問題があったが，定型文を用いた場合と同様，入力文字数が300文字のときはFRRが4.62%，FARが5.64%となり入力文字数が50文字のときに比べて認証精度を大幅に改善することができた．

続いて以下で提案手法の特徴についてさらに詳細な考察を行う．

- 非定型文に関する考察

キーストロークを利用した多くの従来手法では，非定型文を用いた場合の結果は定型文を用いた場合の結果に比べて大幅に悪化したが，提案手法では図5より，大幅な悪化がなかったことで提案手法の有効性が確認できた．

次に，定型文と非定型文のプロファイルデータの違いについて考える．定型文のプロファイルデータには，入力データの全ての二連字が含まれており，その特徴を比較することができる．また，保持しているD_iの値は少なくとも4個以上のデータの平均値であり，非定型文のプロファイルデータに比べると信頼性が高いと考えられる．つまり，プロファイルデータが数多くの二連字データを保持し，かつ数多くのデータの平均値としてそのデータの信頼性を向上させることができれば，非定型文を利用したユーザ認証においても定型文を利用した場合とほぼ同等な結果を得ることができると考えられる．

- 入力文字数に関する考察

今回の実験では300，200，100，50文字と入力文字数を変更して実験を行ったが，この結果から文字を入力する場合，およそ1分ほどの使用で不正使用者を発見することが可能であると考えられる．つまり，この手法を利用することにより不正使用者によるコンピュータの持つ機密情報の改竄を防止することが可能になると考えられる．

- 提案手法の組み合わせ方に関する考察

定型文を用いた場合のord手法だけで識別した結果は，表5より入力文字数が300文字のときはFARが2.00%，FRRが0.44%と良好な結果が得られた．さらに二連字の並べ方を変更することにより，表6の通りFRRの改善を行うことができた．また，この手法により悪化したFARは，さらにabs手法を組み合わせることにより，ord手法単独の結果とほぼ変わらない水準まで改善することができた．

次に，非定型文を用いた場合のord手法だけで識別した結果は，表8より入力文字数が300文字でもFRRが15.38%であったが，二連字の並べ方を変更することにより，表9の通りFRRの改善を行うことができた．また，この手法により悪化したFARは，さらにabs手法を組み合わせることにより，定型文を用いた実験同様，ord手法単独の結果とほぼ変わらない水準まで改善することができた．

- 従来手法との比較に関する考察

文献[8],[13]などで提案されている従来手法では，FRR，FRA共に約10%の結果が得られている．それに対し本稿では，FRRが4.62%，FARが5.64%という結果が得られた．このように認証精度が改善した理由としては，ユーザの特徴として従来手法で主に用いられるabs手法に加えて，キーストロークの速度の順位情報というabs手法とは大きく異なる識別指標を用いたからであると考えられる．表11は，非定型文に対し閾値を$\mu+2\sigma$としてabs手法だけで識別した結果である．2つの手法を組み合わせた結果である表10と比較してFRRの改善が著しいことから，本稿の手法が有効であると考えられる．

表11　abs手法の適用結果(非定型文)

入力文字数	300	200	100	50
本人成功数	55	54	52	52
本人失敗数	10	11	13	13
FRR	12.31%	16.92%	24.62%	33.85%
他人受け入れ数	38	84	84	116
他人拒否数	742	696	696	564
FAR	4.49%	5.51%	6.15%	10.38%

- 継続的な認証システムへの応用に関する考察

今後の課題として，継続的な認証システムを実装する場合，プロファイルデータの更新方法などが考えられる．例えば，プロファイルデータの更新を日常のコンピュータの使用により行なわない場合，ユーザのキータイピング技術の向上や癖の変化に対応できなくなるという問題がある．逆に日常のコンピュータの使用によりプロファイルデータを更新した場合，不正ユーザの長期使用により，プロファイルデータが書き換えられてしまうという問題がある．

また，キーボード操作により作成される文章は日本語だけではないため，英語やプログラミングなど全く異なる文章が入力された場合のユーザ認証精度に関しても実験を行なう必要があると考えられる．

5. むすび

本稿では，ユーザの通常使用時におけるキーストロークを利用した継続的なユーザ認証に不可欠である，非定型文を利用した場合にも高精度な個人認証を行なえる手法を提案した．

また，二連字のD_iの値を昇順で並べ，その並び方に着目することにより，従来研究では難しかった多人数への対応に関して改善することができた．また，ord手法で失敗したデータをabs手法で補うことによって，認証精度の低かった異なる文章を用いたユーザ認証においても精度が向上した．その結果，非定型文でも入力文字が300文字程度あれば，FRRが4.62%，他人拒否率が5.64%となり，提案手法の有効性が確認できた．

今回の実験では，キータイピングを行う被験者のレベルに関係なく閾値を設定したが，文献[13]のようにキータイピングを行う被験者のレベルに合わせて閾値を設定すればさらに正確な認証が可能になると考えられる．また，今回提案した手法では

機密情報の改竄の防止を目的としたため，認証するユーザを 10 人前後として実験をおこなったが，他の目的に応用するためにも，今後さらに多人数を想定した認証実験を行なう必要があると考えられる．

今後の課題としては，今回の実験では入力文字数の減少により，FAR において大幅な精度の低下が見られたが，他の何らかの指標をさらに組み合わせることにより，FAR を改善する必要があると考えられる．また，日本語の文法にも着目することで，同じ二連字でも異なる二連字と捉えることができると考えられる．

さらに，ログイン後のキーボード操作だけでなく，マウス操作も併せて着目することによってキーボード操作を伴わないコンピュータの使用についてもユーザ認証が可能になると考えられる．

謝　辞

元大阪府立大学工学研究科の猪飼武夫先生に，研究を通じて有益な御助言を頂きました．ここに心からの感謝の意を表します．

文　献

[1] J. Ashbourn, Advanced Identity Verification. The Complete Guide., Springer-Verlag, London, 2000.

[2] 瀬戸洋一，“バイオメトリクスを用いた本人認証技術，” 計測制御, vol.37, no.6, pp.395-401, Jun. 1998.

[3] 内田薫，“指紋による個人認証の最前線，” 映情学誌, vol.55, no.2, pp.176-179, Feb. 2001.

[4] 川崎雅也,「網膜」の識別でセキュリテイを守る，月刊エレクトロニクス 株式会社オーム社，東京，1998

[5] 山田浩二，松本弘之，松本勉，“指紋照合装置は人工指を受け入れるか（その３），”2001 年暗号と情報セキュリティシンポジウム予稿集, vol.2, pp.719-724, Jan. 1973.

[6] 小越 康宏，日名田 明，広瀬 貞樹，木村 春彦，“打鍵間時間を基にした認証システムのリズム打鍵による改善，” 情処学論, vol.44, no.02, pp.397-402, Feb. 2003.

[7] 粘川 正充，角田 博保，森 裕子，“アルペジオ打鍵列を利用した個人認証手法の提案，” 情処学論, vol.34, no.05, pp.1198-1205, May. 1993.

[8] 倉橋 勇気，横山 和也，小松 尚久，“キーストロークダイナミクスの特徴と個人照合アルゴリズムの提案，” 個学技法, vol.105, no.40, pp.7-12, May. 2005.

[9] F. Bergadano, D. Gunetti, and C. Picardi, “User Authentication through Keystroke Dynamics,”ACM Trans. Information and System Security, vol.5, no.4, pp.367-397, Apr. 2002.

[10] 泉 正夫，長尾 若，宮本 貴朗，福永 邦雄，“マウス操作 の特徴を用いた個人識別システム，” 個学論 (B), vol.J87-B, no.2, pp.305-308, Feb. 2004.

[11] 樋口正憲，田中敏幸，“位相限定相関法を用いた静脈パターンによる個人認証，” 計測制御, vol.40, no.3, pp.364-366, Mar. 2004.

[12] M. Kendall, J. D. Gibbons, Rank Correlation Methods., Hafner Publishing Company, New York, 1955.

[13] 佐村 敏治，西村 治彦，“キーストロークダイナミックスによる日本語文での個人識別” システム制御情報学会研究発表講演会, 5F3-6, pp.707-708, May. 2007.