氏名入力時における打鍵時間間隔の特徴を利用した個人認証

小宮崚輔[†1]　櫻田ユカリ[†1]　中國真教[†2†3]

概要：指紋や静脈などの生体の特徴を計測する個人認証方式が存在するが，多くの場合，付加的な装置が必要となり，それを装備するためにはコストが増加するなどの課題がある．本研究では，多くのパソコンにおいて標準的に装備されているキーボードを用い，その打鍵における時間間隔の特徴を用いて個人認証を行う方式を検討した．本方式で入力する文字列は，その入力者の氏名としている．氏名は入力する機会が常にあり，入力に慣れていると考えられる．そのため，常に同じリズムで入力できる可能性があり，これを個人認証に利用できる可能性がある．本報告では，打鍵時間間隔を解析した結果と，本認証方式の実用化の可能性について述べる．

キーワード：セキュリティ，アクセス制御，利用者認証，個人認証，打鍵時間間隔

A Method of Personal Authentication by Key Stroke Timing of Full Name Input

RYOUSUKE KOMIYA[†1]　YUKARI SAKURADA[†1]　MASANORI NAKAKUNI[†2†3]

Abstract: A personal authentication method to measure the characteristic of the living bodies such as a fingerprint or the vein exists, however an additional device is necessary, and there is a problem of cost increasing to be equipped with it in many cases. In this study, I examined a method of personal authentication using the characteristic at the interval at time in the key stroke with a keyboard equipped with in many PCs normally. The input character string to use for personal authentication is full name of the person. When people use a PC, they may be used to the input because an opportunity to input it has many. Therefore people may always input with the same rhythm, and personal authentication may use this. This may be available for personal authentication if people can always input with the same rhythm. In this report, described about the result that analyzed interval of key stroke time and the practical use of this method.

Keywords: security, access control, user authentication, personal authentication, key stroke timing

1. はじめに

第三者によるコンピュータの不正利用や個人情報の漏洩を防ぐ手法として，パスワードなどを用いた利用者認証手法が長きに渡り用いられてきた．パスワードによる認証は手軽であり導入コストが安価である反面，パスワードを知ることができれば誰もがその利用者に成りすますことができるという欠点があり，パスワードが漏洩した場合は不正利用を阻止することが困難となる場合が多い．この課題を解決するため，生体の特徴を計測するバイオメトリクス認証が考案され，パソコンやスマートフォンにも搭載されるようになり，認証に関わる安全性は高まっているが，それを実装するためのコストの増大，指紋などの複製による不正な個人認証の発生など，課題は多く残されている．このような課題を解決するためには，複製による再現が困難であり，標準的な装備だけで個人認証を行える方法を検討する必要がある．本研究では，その点に焦点をあて，パソコンを利用する際のキーの入力の様子を解析することで個人認証を行う方法について検討した．

2. キー入力の特徴に着目した個人認証

キー入力に着目した個人認証方式は古くから研究が行われている[1]．しかし，実際の運用に活用できるほどの認証精度を有する方式は少ない[2-3]．過去には製品化された個人認証システムが存在したが，上市としての色合いが濃く，現在は同様の製品は市販されていない．このように，キー入力の解析による個人認証方式の提案は少なかったが，ここ数年の間はこれに関連する研究報告が増え，キー入力の解析による個人認証方式への関心が高まっている．しかし，多くの方式は，キー入力の様子（癖など）を収集し，それを個人認証に用いるために，大量のキー入力が必要である場合が多く，利用者へ負担がかかる上に，大量のキー入力が完了するまで個人認証ができないという欠点がある．そのため，キー入力の特徴を用いた個人認証手法の提案においては，このような課題を解決することも重要である．本研究では，この点も考慮し，多くの課題を解決できる解析方法を検討した．

†1 福岡大学工学部電子情報工学科
Dept. of Electronics Engineering and Computer Science, Fukuoka University
†2 福岡大学総合情報処理センター
Information Technology Center, Fukuoka University
†3 ネバダ大学ラスベガス校
Department of Computer Science, University of Nevada, Las Vegas

3. 打鍵時間間隔の特徴を解析する個人認証

3.1 打鍵時間間隔

本研究では，利用者のキー入力の特徴を表わす情報として打鍵時間間隔を用いている．打鍵時間間隔とは，図 1 および図 2 のように定義している．

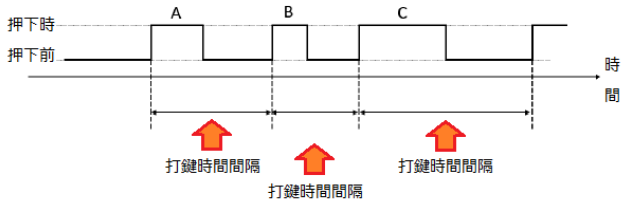

図 1　打鍵時間間隔（通常入力時）
Figure 1　Key stroke timing (nomal)

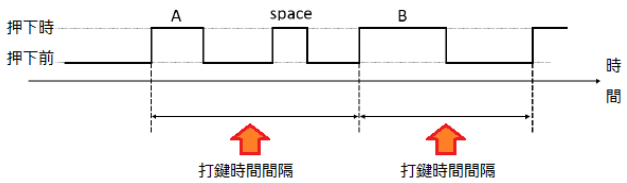

図 2　打鍵時間間隔（スペースキー入力時）
Figure 2　Key stroke timing (with space key)

図 1 は，アルファベットキーが押された時のみ（通常入力時）の打鍵時間間隔を示したものである．各キーが押された瞬間の時間の間隔を打鍵時間間隔と定義している．図 2 は，スペースキーの入力が含まれる打鍵時間間隔の定義である．欧米人のように姓と名の間にスペースキーが入力されたとしても，スペースキーの入力のタイミングは無視する．

3.2 解析対象となる文字列

解析対象となるキー入力の文字列は，その利用者の氏名としている．氏名に着目した理由は，入力に慣れている文字列であれば，入力タイミング（リズム）が入力の度に大きく異なる可能性が低い，という仮説を立てたためである．鉛筆やボールペンを用いて筆記する文字の場合，書き慣れた文字であれば躊躇することなくスムーズに筆記が可能である．逆に，書き慣れない文字であれば，筆記の際に慎重に書くことで緊張を伴い，何度も同じ筆運びで筆記することができない可能性がある．キー入力タイミングにも同じことが言える可能性があり，入力に慣れた文字列であれば，何度入力しても同じリズムで入力できる可能性があると考える．特に氏名については，入力する機会が多くあり，長年に渡って入力しているため，入力に慣れていることが予想される．

3.3 キー配列の違いによる個人認証への影響

氏名のキー入力の打鍵時間間隔を解析する個人認証を検討した理由がもう一つある．それは，多くの氏名の場合，「A から Z までのアルファベットキーしか使わない」という点に深く関係する．キー入力の打鍵時間間隔を解析する際，特殊記号の入力までも解析の対象とした場合，キー入力のリズムを毎回再現できない可能性がある．つまり，個人認証の結果に大きな影響を及ぼすという可能性である．特殊記号のキーの位置は，そのキーボードが採用するキー配列に応じて異なるため，使い慣れていないキー配列のキーボードではリズムを再現できない可能性がある．その上，例え，同じキー配列規格のキーボードであっても，製品毎に特殊記号のキー配置が異なるため，特殊記号が含まれる文字列を入力する際には，リズムを再現することが困難と考えられる．本研究では，パソコンを買い換えることでキー配列が変わった場合でも，他人のパソコンを借りた場合でも，リズムを再現できることが重要と考え，解析対象となる文字列に含まれる文字はアルファベットのみとし，アルファベットのみで構成されることが多い「氏名」を解析対象とした．

4. 個人認証の実験

4.1 実験内容

今回の実験では，個人認証システムのプロトタイプの開発を目指し，3 名の被験者による小規模な実験を行った．各被験者のパソコンの使用歴を次に示す．

【被験者 A】
パソコン使用歴は 10 年未満．理工系の大学生．
【被験者 B】
パソコン使用歴は 10 年未満．理工系の大学生．
【被験者 B】
パソコン使用歴は 20 年以上．業務においてパソコンを頻繁に使用．ソフトウェア開発などの経験もある熟練者．

このような被験者によって次に示す 2 つの実験を行った．

【実験 1】
3 人のうちの 1 人を正規利用者，他 2 人を不正利用者と仮定した個人認証実験．
【実験 2】
氏名は同一であるが互いに別人である利用者が複数存在すると仮定し，それぞれの利用者を判別する実験．

4.2 実験環境

プロトタイプの開発並びに個人認証の実験（キー入力情報の取得）を行った主なハードウェア環境，そして，ソフトウェア環境について，表 1 に示す.

表 1 実験環境
Table 1 Experiment environment

項目	内容
PC 型番	TOSHIBA dynabook N40/TG
CPU	Intel Atom(TM) 1.44GHz
メモリ	2.00GByte
OS	Windows 10 home 64bit
開発言語	Hot Soup Processor (HSP) Ver. 3.4

本認証方式は身体的特徴を用いる個人認証ではないため，表 1 に示したハードウェア以外の付加的なハードウェアを用いていない.

4.3 打鍵時間間隔の解析方式

前述の 2 つの実験を行う際，複数の解析方式によってキー入力を解析した．それぞれの方式を次に示す．これらの方式における「入力データ」とは，実際に個人認証を行う際に採取したデータを指し，「サンプルデータ」とは，個人認証を行うために予め登録した比較対象となるデータを指し，入力データと比較するためのものである.

【方式 1】
入力データと、サンプルデータの算術平均値の差をサンプルデータの標準偏差と比較.
【方式 2】
入力データと、サンプルデータの算術平均値のユークリッド距離をサンプルデータの標準偏差の総和と比較.
【方式 3】
入力データと、サンプルデータの算術平均値のマンハッタン距離をサンプルデータの標準偏差の総和と比較.
【方式 4】
入力データと、サンプルデータの算術平均値のコサイン類似度を算出し，閾値と比較.

次に，解析方式の詳細について説明する．その説明のために，次の記号を使用する.

time(len, n)： サンプルデータ取得時の打鍵時間間隔
ave_{len}： サンプルデータの算術平均値
S_{len}： サンプルデータの標準偏差
ただし，*len*，*n*（≧0）は以下のものとする.
len： 取得した打鍵時間間隔の総数
n： サンプルデータの入力回数及び番号（0～9）

そして，サンプルデータから作成するモデルを次に示す.

$$ave_{len} = \frac{1}{10}\sum_{k=0}^{9} time(len,k) \quad (1)$$

$$s_{len} = \sqrt{\frac{1}{10}\sum_{k=0}^{9}\{ave_{len} - time(len,k)\}^2} \quad (2)$$

4.4 方式 1 による解析

この方式では，標準偏差を閾値とする．氏名入力時の打鍵間隔時間と打鍵間隔時間の算術平均値の差の絶対値を求め，標準偏差と比較する．このとき，すべての差が標準偏差以下に収まれば，比較したデータは同一人物のものと見なし，認証する．次に，認証に用いる条件式を示す.

$$\forall_{len}\{|ave_{len} - time(len,n)| \le s_{len}\} \quad (3)$$

ただし，*n* は入力データとして認証に使用したサンプルデータの番号である.

4.5 方式 2 による解析

この方式では，取得した打鍵間隔時間データおよびモデルを *n* 次元ベクトルとし，(4) 式より，氏名入力時の打鍵間隔時間と打鍵間隔時間の算術平均値のユークリッド距離 *d* を用いて認証を行う．打鍵間隔時間の標準偏差の総和を閾値とし，(5) 式に示すように求めた距離が閾値以下であれば認証する.

$$d = \sqrt{\sum_{k=0}^{len-1}\{ave_k - time(k,n)\}^2} \quad (4)$$

$$d \le \sum_{k=0}^{len-1} s_k \quad (5)$$

4.6 方式 3 による解析

この方式は，距離の求め方によって認証結果にどのような違いが出るかを確認するため実装した. 前述の【方式 2】と同様に, 打鍵間隔時間の標準偏差の総和を閾値とし, (6) 式より，氏名入力時の打鍵間隔時間と打鍵間隔時間の算術平均値のマンハッタン距離 d を算出し，(5) 式に示すように求めた距離が閾値以下であれば認証する.

$$d = \sum_{k=0}^{len-1} |ave_k - time(k,n)| \quad (6)$$

4.7 方式 4 による解析

この方式では，任意のコサイン類似度[4]を閾値とする．ここで，コサイン類似度とは，2 ベクトル *x*, *y* のなす角の余弦 *cos* のことをいい，ベクトルの向きの近さを類似性の指標としたものである．コサイン類似度は向きが一致すれば 1，直交していれば 0，反対を向いていれば-1 となる．また，この方式では，閾値として任意のコサイン類似度を設定し，認証に用いるコサイン類似度の算出は打鍵間隔時間と打鍵間隔時間の算術平均値を用いて行うため，打鍵間隔時間の標準偏差は使用しない．

(7) 式より，氏名入力時の打鍵間隔時間と打鍵間隔時間の算術平均値のコサイン類似度を求め，その類似度が設定した閾値 *SIM* 以上であれば認証する．

$$sim = \frac{ave \cdot data(n)}{\|ave\| * \|data(n)\|} \quad (7)$$

$$sim \geq SIM \quad (8)$$

ただし，*data*(*n*), *ave* は打鍵間隔時間と打鍵間隔時間の算術平均値の *n* 次元ベクトルであり，*n* は入力データとして認証に使用したサンプルデータの番号である．

5. 各認証方式による評価結果

実験 1，実験 2 の結果をそれぞれ表 2，表 3 に示す．得られた結果より，それぞれの実験での各認証方式の評価を行い，最後に 2 つの実験を通して得られた各認証方式の評価を行う．

5.1 実験 1 の結果

結果が最も良好といえるものは，【方式 3】のマンハッタン距離を用いた認証方式であり，試行回数が少ないとはいえ，FAR（False Acceptance Rate：他人受入率），FRR（False Rejection Rate：本人拒否率）共に 0%と理想的な結果となった．しかし，距離の算出法が異なる【方式 2】では理想的な結果とはならず，FRR は 0%であったが，FAR が 41.7%と非常に高い値となった．また，【方式 1】における単純にデータの差を求める認証方式は，FRR が 100%と正規利用者を完全に拒否する結果となった．さらに，【方式 4】のコサイン類似度を用いた認証方式では，FAR は低かったが FRR が十分な値ではなかった．

5.2 実験 2 の結果

【方式 1】，【方式 3】と【方式 4】においては，閾値のコサイン類似度を 0.97 以上とした場合の FAR は 0%となった．しかし，【方式 2】では，FAR が 75%と非常に高い値となった．

表 2　実験 1 の結果
Table 2　Result of experiment 1

解析方式	閾値		FAR	FRR
方式 1	各標準偏差		0.00%（0/60）	100.00%（30/30）
方式 2	標準偏差の総和		41.70%（25/60）	0.00%（0/30）
方式 3	標準偏差の総和		0.00%（0/60）	0.00%（0/30）
方式 4	任意のコサイン類似度	0.950	13.30%（8/60）	3.33%（1/30）
		0.970	1.67%（1/60）	6.67%（2/30）
		0.975	0.00%（0/60）	13.30%（4/30）
		0.980	0.00%（0/60）	20.00%（6/30）

表 3　実験 2 の結果
Table 3　Result of experiment 2

解析方式	閾値		FAR	FRR
方式 1	各標準偏差		0.00%（0/60）	
方式 2	標準偏差の総和		75.00%（45/60）	
方式 3	標準偏差の総和		0.00%（0/60）	
方式 4	任意のコサイン類似度	0.950	18.30%（11/60）	
		0.970	0.00%（0/60）	
		0.975	0.00%（0/60）	
		0.980	0.00%（0/60）	

5.3 各認証方式の評価

実験 1 および実験 2 の結果から，今回最も優れていた認証方式は【方式 3】であったことが分かった．この認証方式は，すべての実験において理想的な結果が得られた．また，【方式 4】は，【方式 3】には劣るものの，十分な結果を得られたと判断する．さらに，これら 2 つの認証方式と組み合わせれば，FRR の値は低く問題はなかったが，FAR の値が非常に高かった【方式 2】も個人認証システムとして利用できる可能性がある．しかし，【方式 1】は FRR が 100%であり，登録者のすべての入力を拒否してしまい，個人認証システムとして利用することは極めて困難と判断する．

6. 考察

本研究では利用者の氏名入力時のキー入力の時間的特徴から個人認証を行うシステムについて開発を進めてきた．ここでは，開発した個人認証システムのプロトタイプに関する考察について述べる．まず，開発した個人認証システムに実装した 4 つの異なる解析方式それぞれについて以下にまとめる．

【方式 1】
入力データと、サンプルデータの算術平均値の差をサンプルデータの標準偏差と比較

・個人認証システムとしての利用は困難
この解析方式では FRR が 100%となり，正規利用者の入力をすべて拒否してしまい，個人認証システムとして利用が困難と判断する．

【方式 2】
入力データと、サンプルデータの算術平均値のユークリッド距離をサンプルデータの標準偏差の総和と比較

・単体での利用の場合は精度が不十分
この解析方式では FRR は 0%と理想的な値となったが，FAR の値，特に正規利用者と同姓同名の利用者が誤って認証される割合が 75%と非常に高く，この解析方式単体での利用には不十分である．

・他の解析方式と組み合わせによる精度向上の見込み
FRR の値は理想的であることから，FAR の値が低かった【方式 3】,【方式 4】と組み合わせることで，個人認証システムとして利用できる可能性が高まる．

【方式 3】
入力データと、サンプルデータの算術平均値のマンハッタン距離をサンプルデータの標準偏差の総和と比較

・認証システムとして理想的な結果を示した
この認証方式では FAR，FRR の値は共に 0%であり，認証システムとして理想的な結果を示した．

【方式 4】
入力データと、サンプルデータの算術平均値のコサイン類似度を算出し，閾値と比較

・閾値として高いコサイン類似度を設定する必要性
正規利用者以外の入力であっても，算出されるコサイン類似度は高く，閾値の値を高めに設定する必要があった．

・サンプルデータのばらつきの影響が大きい
正規利用者のサンプルデータのばらつきが大きいと，他の入力に比べ算出されるコサイン類似度が若干低くなってしまうことから，閾値を高く設定した場合の FRR の増加につながったと考えられる．

次に，本手法の利点について以下にまとめる．

・成りすましの発見が可能
開発した個人認証システムでは，実験 1，実験 2 の結果から，正規利用者本人の氏名を知っている者による入力があった場合，または，正規利用者によって登録されている文字列が漏洩した場合でも，氏名入力時の時間的特徴から 2 人が別人であるということを判別可能である．

・付加的な装置が不要
開発した個人認証システムは，キーボードによるキー入力から認証を行うことから，付加的な装置を必要としない．したがって，本手法の導入は比較的容易と言える．

・正規利用者によって登録される文字列は変更する必要性が低い
本手法では，正規利用者には自身の氏名を入力してもらうことから，登録文字列を思い出すことができないことで認証ができないという事態を招く可能性は非常に低い．また，登録文字列が漏洩した際も他者が不正に認証される可能性が低いため，登録文字列を変更する必要性は低いと考えられる．

最後に，本手法の問題点について以下にまとめる．

・利用者自身による利用における配慮の必要性
時間的特徴により認証を行っているため，氏名入力時のキーストロークを真似られると，他人が不正に認証される恐れがある．そのため，正規利用者は氏名入力を常に一定間隔の打鍵リズムでキー入力を行わないよう配慮する必要がある．

・実験データの不足
本研究において実験を行ったのは 3 人のみであることから，被験者をさらに増やした実験が必要と判断する．また，今回の 3 人の被験者のパソコンの利用頻度がほぼ等しく，利用頻度が異なる人々のデータでの実験も必要と判断する．

参考文献

[1] 中國真教，堂薗浩，野口義夫：キーボード入力の監視による不正利用者の判別方法，情報処理学会論文誌，No.41 Vol.12，pp. 3276-3284 (2000).
[2] Hiroshi Dozono, Hisao Tokushima, Masanori Nakakuni and Yoshio Noguchi：The Analysis of Key Stroke Timings Using Self Organizing Maps and its Application to Authentication, The 2006 International Conference on Security and Management (SAM ’06), pp.100-105, (2006.06).
[3] 小松篤史，三好力：自己組織化マップを使用したキーボード入力による個人認証，情報処理学会第 74 回全国大会 3D-3 2 pp.81-82, (2012).
[4]データ分析・マイニングの世界 by SAS，距離・類似度について：http://wikiwiki.jp/cattail/.