共同研究報告書

（研究題目）「バイオメトリクスセキュリティ評価基準に関する研究」

（研究内容）１．バイオメトリクスにおける脅威・脆弱性の明確化
２．脆弱性の程度を測る評価実験の実施（顔、音声、署名）

平成 17 年２月１５日

早稲田大学理工学部　教授　小松　尚久

目次

Ⅰ．オンライン署名照合の脆弱性に関する検討

１．目的

本稿では，バイオメトリクスの一つである筆記情報を用いたバイオメトリクス認証技術（筆者認識技術[1]）に着目し，特に，オンライン署名照合システム，ならびにテキスト独立型筆者照合システムにおいて，テンプレートや照合結果から元の生体情報が推定できる脆弱性（以下，推定に関する脆弱性）[2]について検討した結果を報告する．

２．脆弱性評価の概要

表 1 に示すように，筆者認識システムは複数の観点から分類されるが，本稿では，テキスト依存型オンライン筆者認識システムの一例として，オンライン署名照合システム，ならびにオンラインテキスト独立型筆者照合システムを対象とする脆弱性評価を行った．

表 1：筆者認識システムの分類

認識の種別	照合，識別
テキストへの依存性	テキスト依存，テキスト独立，テキスト提示
筆記情報の性質	オンライン，オフライン

また，本稿では，推定に関する脆弱性に着目したが，脆弱性評価を行うにあたり，この脆弱性を利用したいかなる脅威が存在するかを明確にする必要がある．現在，推定に関する脆弱性を利用した脅威としては，推定した生体情報によるなりすましが知られており，具体的な方法として，

① バイオメトリクス装置に保管された正当な利用者のテンプレートから生体情報を推定し，推定した生体情報をバイオメトリクスシステムに入力することでなりすましを行う方法．
② 正当な利用者の照合結果（類似度や距離）に漸近するように生体情報を徐々に変化させ，繰り返し照合を行うことで，なりすましを行う方法（ヒルクライミング・アタック）．

が挙げられる[2]．そこで本稿では，オンライン署名照合システムを対象としたヒルクライミング・アタック，ならびにテキスト独立型筆者照合システムを対象としたテンプレートからの筆記情報の推定によるなりすましについて検討を行った．

３．ヒルクライミング・アタックによる偽筆生成となりすまし

図 1 は，ヒルクライミング・アタックを用いたオンライン署名照合システムにおける具体的ななりすましの一手法を示したものである．同図において，署名照合システムが，入力された署名情報とテンプレートとの距離，あるいは類似度を照合結果として出力する場合，正当な利用者の照合結果に漸近するように署名情報を逐次的に変化させて偽筆を生成し，繰り返し照合を行うことによりなりすましを行う脅威が考えられる．本章では，このようなヒルクライミング・アタックに基づき，計算機上で生成された偽筆により署名情報を推定する方法について述べる．

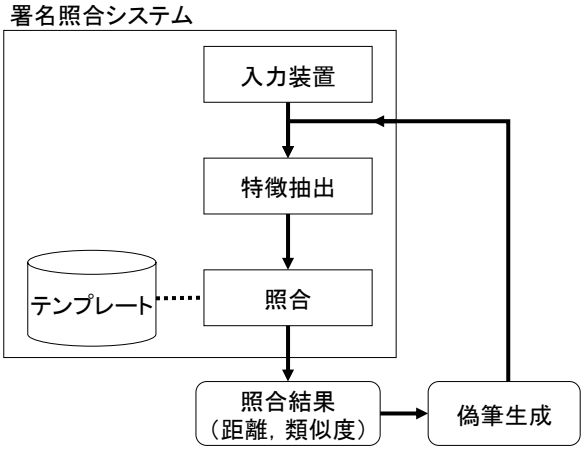


図 1：ヒルクライミング・アタックによる偽筆生成

３．１　偽筆生成手順

本節では，上述のヒルクライミング・アタックを用いた偽筆生成の手順について述べる．

【step 1】参照テーブルの作成

様々な筆記文字を対象として，筆記データの筆記方向に関する時系列データを量子化し，その出現確率を算出した参照テーブルを作成する．まず，入力装置から入力された筆記データの標本点における時系列データを $A^* = a_1^*, a_2^*, \Lambda\ , a_I^*$ と表す．ここで，I は時系列データのサンプル数を表し，$a_i^* = (x_i^*, y_i^*)$ とする．このとき，連続する 2 つの標本点を結んだ線分の方向を，図 2 に示すように L 値に量子化し，量子化後のデータ系列を $A = q_1, q_2, \Lambda\ , q_{I-1} \in \{1,2,\Lambda\ , L\}$ と定義する．また，$q_i = m$ における $q_{i+1} = n$ の条件付確率 $P_m(n)$ を求める．参照テーブルには，$P_m(n)$ および $q_i = m$ ，$q_{i+1} = n$ を満足する標本点 a_{i+1}^*, a_{i+2}^* の間の x, y の平均距離 $\overline{x_{mn}} = \frac{1}{l}\sum(x_{i+2}^* - x_{i+1}^*)$，$\overline{y_{mn}} = \frac{1}{l}\sum(y_{i+2}^* - y_{i+1}^*)$ を保存する．ここで，l は $q_i = m$ の下で $q_{i+1} = n$ となる標本点の数を表す．

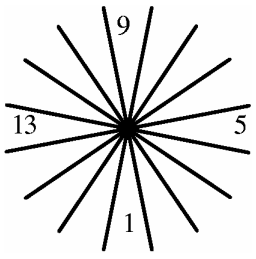


図 2：L=16 のときの量子化方向

【step 2】データの変更箇所の選択

偽筆データ $C = c_1, c_2, \Lambda\ , c_k, \Lambda\ , c_K$ に対して変更を加える部分 k を，データ系列の先頭から順に選択する．ここで，K は偽筆データのサンプル数を表す．

【step 3-1】データの変更(1)

参照テーブルに基づき出現確率の高い方向 n を選択して偽筆データを変更する．選択された $c_k=(x_k,y_k)$ に対し，変更後のデータを $c_k^*=(x_k^*,y_k^*)$ とするとき，$x_k^*=x_{k-1}+\overline{x_{mn}}$，$y_k^*=y_{k-1}+\overline{y_{mn}}$ となるように変更を加える．

【step 3-2】データの変更(2)

出現確率の高い方向から順に L 方向全てに対して変更を加え，照合時におけるテンプレートと偽筆データの距離 D が減少しない場合，c_k および c_{k-1} の関係が現時点で最適であると仮定し，$c_{k+1}=(x_{k+1},y_{k+1})$ に対する変更後のデータを $c_{k+1}^{**}=(x_{k+1}^{**},y_{k+1}^{**})$ とし，$x_{k+1}^{**}=x_k^*+(x_k-x_{k-1})$，$y_{k+1}^{**}=y_k^*+(y_k-y_{k-1})$ となるように変更を加える．

【step 4】変更データの保存

距離 D が減少した場合，変更した偽筆データ（c_k^* または c_{k+1}^{**}）を保存して【step 2】へ戻る．減少しない場合，変更前の状態（c_k または c_{k+1}）に戻して【step 3-1】または【step 3-2】へ戻る．

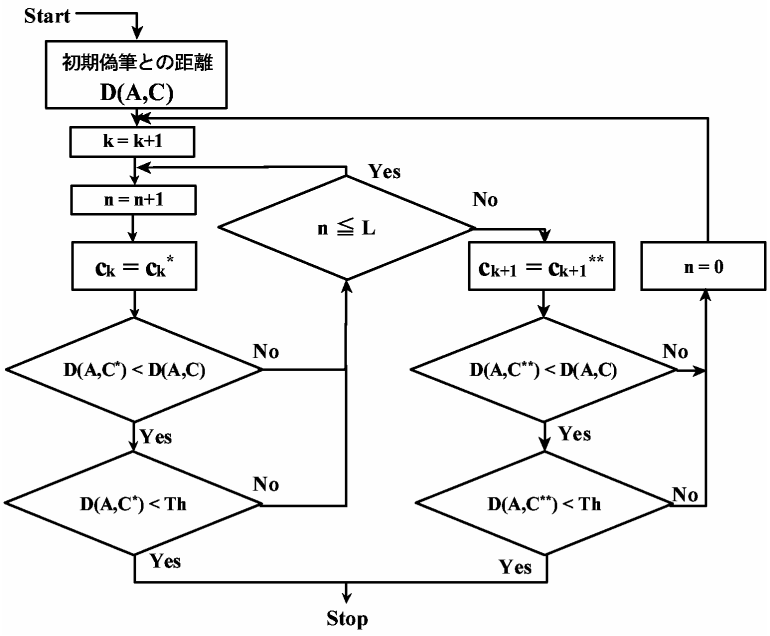


図 3：偽筆生成アルゴリズム

３．２　偽筆生成実験

ヒルクライミング・アタックによる偽筆生成を利用したなりすましの脅威を検証するため，実際の筆記データを使用したシミュレーション実験（偽筆生成実験）を行い，署名照合システムの脆弱性を評価した．以下，実験の概要と実験結果について述べる．

3．2．1　実験の概要

(a) 署名照合アルゴリズム

署名照合システムの一例として，文献[3]の DTW（Dynamic Time Warping）を利用した署名照合アルゴリズムに基づき，偽筆生成実験を行った．実験では，署名照合の前処理として，連続して同一の標本値を与える停留点の除去，標本点数や筆記データの位置，大きさに関する正規化を施した．なお，文献[3]では照合時の特徴量に筆圧の値が含まれるが，本実験では，ペンの位置座標と筆記方向のみを特徴量として用いた．

(b) 入力装置（タブレット）の仕様

実験に使用した入力装置の仕様を表 2 に示す．

表 2：入力装置の仕様（WACOM 社，intuos2，I-1220 を使用）

読み取り分解能	0.01mm（最高値）
読み取り精度	±0.25mm
読み取り速度	200 ポイント／秒（最大値）
読み取り可能高さ	6mm（ペン使用時）

(c) 筆記データの諸元

実験に使用する署名は漢字署名とし，筆記時には署名記入欄を設けた用紙をタブレット上に置き，紙の上から入力専用のペンで筆記データを入力した．ここで，署名記入欄の大きさは，代表的なクレジットカード会社の売上票に設けられた署名欄のサイズを参照し，縦 14mm，横 60.5mm とした．また，筆記データについては，5 名（A～E）の被験者よりそれぞれ 20 個の署名を採取し，うち 3 個をテンプレート作成用データ，残りの 17 個を本人確認のための照合用データとした．なお，筆記データの採取にあたっては，1 日 1 回につき 5 個の署名を採取し，これを 4 日連続して行い 20 個の署名を採取した．

(d) 偽筆データの諸元

ヒルクライミング・アタックによる偽筆生成では，アタックの開始時に与える偽筆（初期偽筆）の精巧さにより，なりすましの難易度が変化するものと考えられる．そこで，本実験では，人の手により書かれた以下の 4 種類の偽筆を用意し，初期偽筆として署名照合システムに与えた．

1. 単純偽筆(1)：攻撃者が自分自身の筆記動作および筆跡で登録者の署名を入力したもの．
2. 単純偽筆(2)：攻撃者が登録者の筆跡を閲覧し，筆跡を真似て署名を入力したもの．
3. 模倣偽筆(1)：攻撃者が登録者の筆跡をなぞることで，署名を入力したもの．
4. 模倣偽筆(2)：(攻撃者と登録者の筆順が異なる場合) 攻撃者が登録者の筆順どおりに登録者の筆跡をなぞり，署名を入力したもの．

ここで，上記の偽筆は番号が大きいものほどより精巧であり，攻撃者は登録者の署名に関する情報を事前に多く入手した状態で攻撃を開始することが可能となる．本実験では，各被験者に対し，上記 4 種類の偽筆を 16 個ずつ用意した．

3．2．2　実験結果

ヒルクライミング・アタックの開始時に，署名照合システムが偽筆を受け入れることがない状況を想定し，上記 4 種類の偽筆全てに対して FAR の値が 0 かつ FRR の値が最小となるような照合のしきい値を被験者ごとに設定した．なりすましの成功までに要した偽筆データの変更回数の平均値，最小値，最大値の実験結果を表 3 に示す．また，偽筆データの一例として，単純偽筆(2)を初期偽筆としたときに生成された被験者 A の偽筆データを図 4 に，偽筆データ生成時のデータの変更回数 n と照合時の距離 D の関係を図 5 に示す．

表 3：なりすまし成功までの偽筆データの変更回数

(a) 単純偽筆 (1)

被験者	平均値（回）	最小値（回）	最大値（回）
A	1303	707	2750
B	1739	694	3093
C	994	454	3853
D	1588	702	3711
E	1725	948	5876

(b) 単純偽筆 (2)

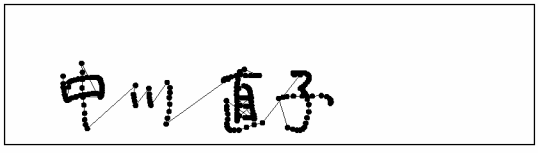
被験者	平均値（回）	最小値（回）	最大値（回）
A	587	169	1830
B	1715	980	3967
C	722	213	1774
D	987	332	2566
E	947	25	2022

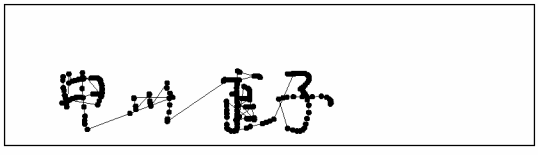
(c) 模倣偽筆 (1)

被験者	平均値（回）	最小値（回）	最大値（回）
A	577	6	917
B	2071	641	6649
C	410	32	625
D	446	145	678
E	690	115	1354

(d) 模倣偽筆 (2)

被験者	平均値（回）	最小値（回）	最大値（回）
B	689	34	3201
C	460	37	808





(a) 初期偽筆　　(b)生成偽筆

図 4：(a)被験者 A の初期偽筆と(b)ヒルクライミング・アタックにより生成された偽筆

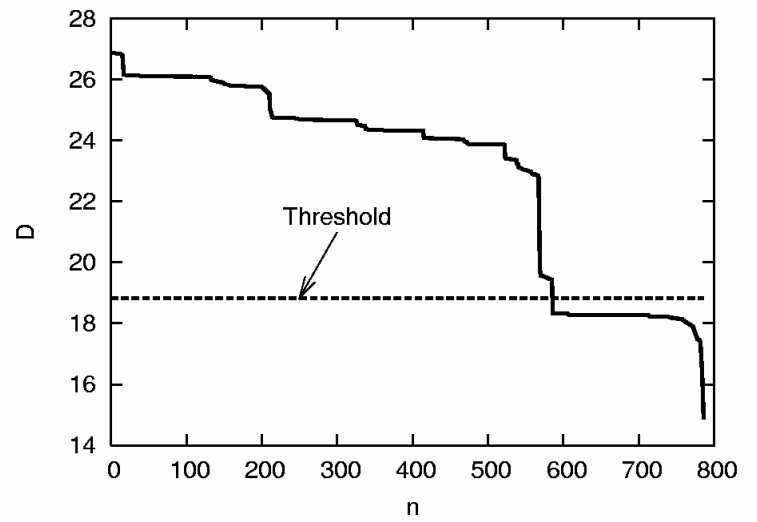


図 5：偽筆データの変更回数 n と照合時の距離 D の関係

実験結果より，なりすましが成功するまでに要する偽筆データの変更回数は，初期偽筆とテンプレートとの距離に依存するものの，どの初期偽筆に対しても有限回の変更により本人として受け入れられる偽筆データが生成されていることが確認される．また，表 3 の(c)や(d)にみられるように，初期偽筆が精巧な場合，攻撃開始時点における偽筆データとテンプレートとの距離は相対的に小さく，数回から数十回のデータの変更で，なりすましが可能となる偽筆が生成される可能性のあることも確認される．これらの結果より，人の手により書かれた偽筆を用いてなりすますことが困難な署名照合システムに対しても，ヒルクライミング・アタックを適用することにより，なりすましが可能となる偽筆を生成できる可能性のあることが確認された．

３．３　まとめ

本章では，オンライン署名照合システムにおける推定に関する脆弱性について検討し，実験の結果，ヒルクライミング・アタックにより，正当な利用者として受理される偽筆を生成できる可能性のあることが確認された．今後の課題として，他の脆弱性の分析ならびに偽筆対策についての検討が残されている．

４．テンプレート情報を用いた筆記情報の推定となりすまし

図 6 は，テキスト独立型筆者照合システムにおける合成偽筆による詐称の一手法を示したものである．同図において，テキスト独立型筆者照合システムは，登録時には，入力された筆記情報をストローク(一画)毎に分割し，クラスタリング処理を施すことで複数のカテゴリに分類，各カテゴリについて，複数の筆者に共通する共通モデル Co-HMM と個々のユーザモデル User-HMM を生成する．そして照合時には，それぞれの HMM に対する尤度を算出し，最も尤度の高いカテゴリについて対数尤度比を算出し，閾値判定を行う．本章では，このようなテキスト独立型筆者照合システムに対し，テンプレートから筆記情報を推定し，生成された偽筆によりなりすましを行う方法について述べる．

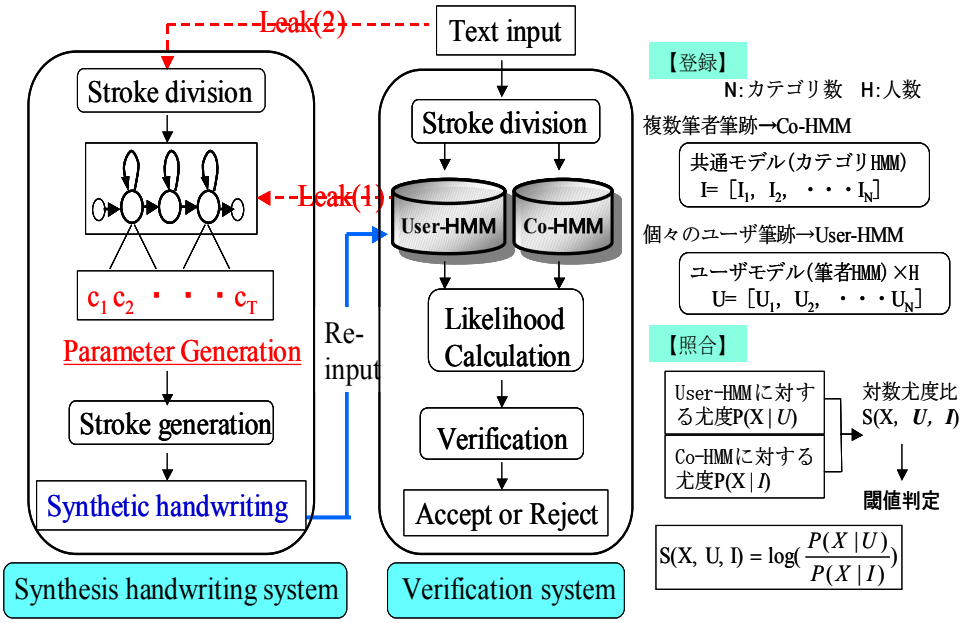


図 6：テキスト独立型筆者照合システムに対する合成偽筆による詐称の流れ

４．１　偽筆生成手順

本節では，HMM テンプレートからの偽筆生成の手順について述べる．

【step 1】詐称用テンプレートの作成

詐称用 HMM には，図 6 に示した次の２種類の漏洩パターンを想定した．

Leak(1)：正規の HMM テンプレート(User-HMM)の漏洩．

Leak(2)：オンラインデータの漏洩を想定し，複数の筆者から正規の Co-HMM とは別に詐称用に生成した Co-HMM に学習し，テンプレートを推定する．

【step 2】状態継続長の決定

HMMには図6のようにLeft-to-Rightモデルを用い，与えられた時間に通過する各状態から偽筆パラメータcを抽出する．その際，時刻 $t = 1 \sim T$ の間に状態 $i = 1 \sim K$ を通過するとし，状態 i が d_i 回継続する分布列の決定には，文献[4]の状態継続長の決定に基づき，

$$d_i = m_i + \rho\sigma_i^2$$
$$\rho = (T - \sum_{k=1}^{K} m_i) / \sum_{k=1}^{K} \sigma_i^2$$

と定める．但し，m_i と σ_i^2 は，それぞれ，状態 i に関するガウス分布の平均と分散である．

【step 3】パラメータの生成

状態継続長 Q に沿って観測される長さ T の出力ベクトル系列 $O = [o_1, o_2, \cdots, o_T]$ を，静的特徴 c_t および動的特徴 Δc_t から求める． c_t は各状態における平均ベクトルを示し，一つの状態が継続している間は一定値をとるが，別の状態に遷移する際に生じる不連続な変化を補うために，遷移確率 a_t を用い， $\Delta c_t = a_t c_t + (1 - a_t) c_{t+1}$ とする．そして，状態継続長に従い静的特徴と動的特徴を並べたストロークを偽筆ストロークとして入力する．

４．２　テンプレート情報を用いた偽筆による評価実験

テンプレート情報から筆記情報を推定し生成した偽筆によるなりすましの脅威を検証するため，シミュレーション実験（偽筆耐性実験）を行い，テキスト独立型筆者照合システムの脆弱性を評価した．以下，実験の概要と実験結果について述べる．

４．2．１　状態継続長に従い生成した偽筆ストロークの耐性評価

実験に使用する筆記データには，データベース[5]を用い，漢字のみを入力データとして判別した．タブレットの条件は，frame size：60×60 ピクセル（１文字当たり），sampling rate：66 ms であり，ペンの位置座標と筆記方向のみを特徴量として用いた．そして，詐称用テンプレートの生成諸元を表 4 に示す．

表 4：詐称用テンプレート諸元

	詐称用テンプレート(a)	詐称用テンプレート(b)
［カテゴリHMM］ 学習データ	正規HMM生成用と同データ 1000字（50字×20人）（漢字）	正規HMM生成用と別データ 1000字（25字×40人）（漢字）
［筆者HMM］ 個人性学習データ	正規HMM生成用とは同データ 各人200 or 500字(漢字)	正規HMM生成用とは別データ 各人200 or 500字(漢字)

詐称用テンプレート(a)は，正規のテンプレートの漏洩と同条件(500 字)，あるいはそれに近い条件(200 字)を想定し生成した．それに対し，詐称用テンプレート(b)は，カテゴリ HMM，筆者 HMM ともに正規のテンプレートとは異なる条件で生成した．一回の照合に任意の照合データ 5 文字を用いた時の照合結果を図 7，図 8 に示す．

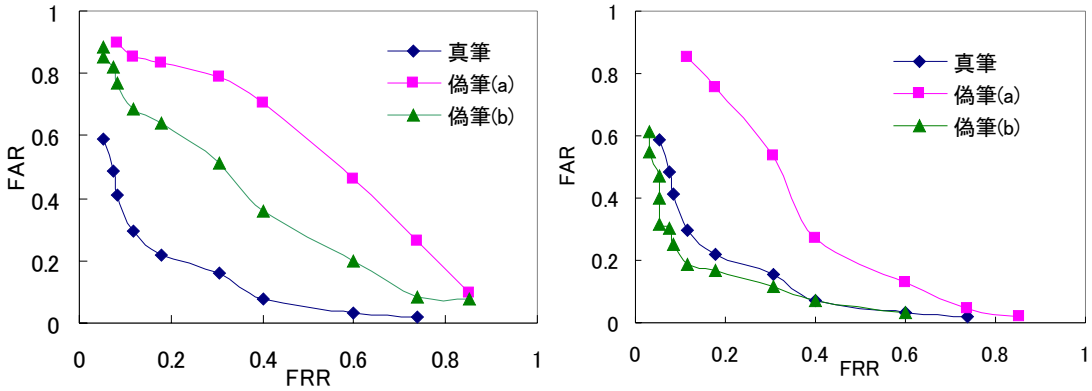


図 7 ：学習量が十分にある場合　　図 8：学習量が少量の場合

偽筆については、正規のテンプレートとほぼ同条件の場合，学習量に限らず高い詐称受理率を示している。また、異なる初期モデル(カテゴリ HMM)を用いた場合，少量の学習量では自然筆跡以下の詐称受理率であるが，十分な学習量を用いると自然筆跡以上の詐称受理率を示している．但し，状態継続長に基づき生成した偽筆ストロークは新規にテンプレートから生成するため，テキスト提示型に対しては文字としての位置や組み合わせを詐称者側で指定しなくてはならない条件が加わる．

4．2．2　既存ストロークへのパラメータ付加を行った偽筆の耐性評価

位置指定を施す必要がないように，詐称者が入力した自然筆跡に対してなりすまし対象者の特徴を付加することで偽筆を生成した．その手順の例を以下に示す．

(1)詐称者が入力した時系列データを図 2 を用い量子化し，$O = \{5,5,8,8,8,2,9,9,10\}$ とする．

(2)同方向ごとに時系列データ O を分類し，$O : |5,5|8,8,8|2|9,9|10|$ とする．

(3)一つのグループを一状態としてみなし，状態系列 $Q : |q_1|q_2|q_3|q_4|q_5|$ とする．

(4)各状態間に詐称用テンプレートから抽出したパラメータを埋め込む．

図 9 にその照合精度特性を示し，図 10 に既存ストロークへの特徴付加を行い生成した偽筆を示す．その結果，自然筆跡に付加するベクトルが多いほど歪みが生じているが，より詐称受理率が上昇する傾向がある．

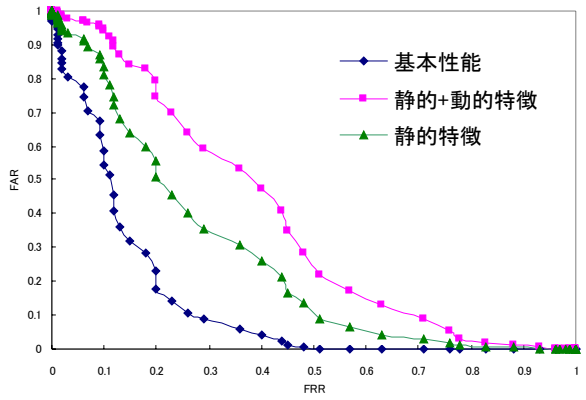
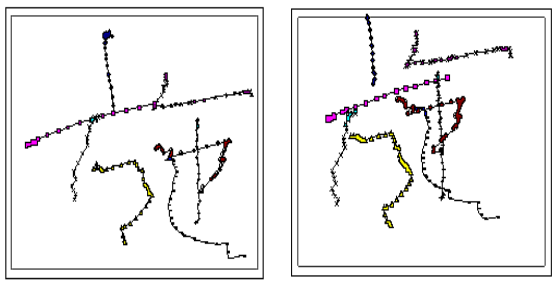


図 9：偽筆の照合精度特性



(a)静的特徴のみ　(b)静的＋動的特徴導入

図 10：既存ストロークに特徴付加した偽筆

４．３　考察

自然筆跡による照合精度特性がEER 約20 %を示す筆者照合システムに対し，他者の自然筆跡の角度変換点に静的特徴の1 ベクトルを付加することで詐称者受理率が向上し、EER は

約36 %に達し照合精度特性が悪化している．さらに，静的特徴と動的特徴の2ベクトルを付加した場合，よりテンプレートの特徴が強調され，詐称者受理率が上昇しEER は約44 %に達した．しかし同時に，加えるベクトル数が多くなることで，文字に歪みが生じた．これは角度変換点に対し付加したパラメータベクトルの位置と方向性が，正しく合致しなかったためと考えられ，筆記情報に動的特徴を付加した位置が特定されてしまう可能性がある．そこで，こうした形状の歪みを避けるには，どの部分にベクトルを埋め込むかを再度検討する必要がある．一方，偽筆対策として埋め込み検知を行うために，前後のベクトルの関連性を検討することが考えられる．

４．４　合成偽筆によるなりすまし対策

合成偽筆によるなりすまし対策として，合成偽筆と自然筆跡を判別する必要がある．通常，書き慣れた署名などは筆者内変動が小さく，詐称者が模倣した自然筆跡は大きくなることが予測される．テキスト独立型の場合，署名のように書き慣れた筆跡に依存しないが，書き慣れたカテゴリストロークが存在する可能性はある．そこで，自然筆跡と合成偽筆による各カテゴリの筆者内変動を検証した．

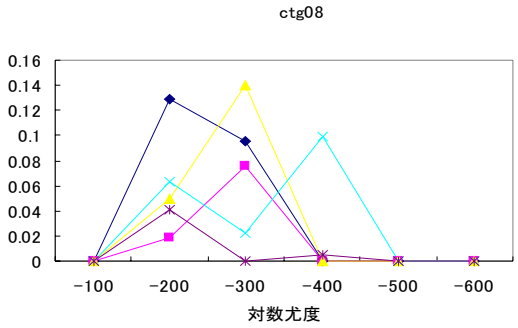


図11: 自然筆跡の筆者内変動

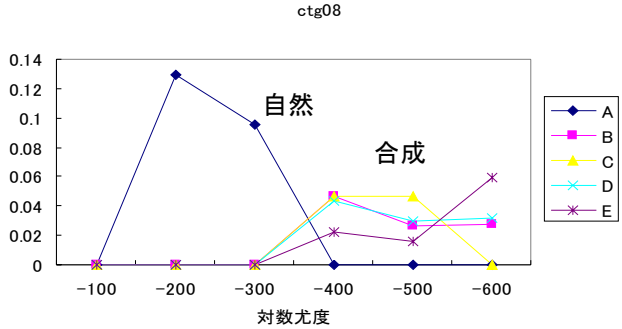


図12: 合成偽筆の筆者内変動

図 11，12 は，あるカテゴリにおいて，筆者 A の筆者 HMM に対して，本人と詐称者 B～E の筆記ストロークとの対数尤度を算出した結果を示す。ここで，100 毎の範囲でストロークを分類して，同カテゴリ内において類似性の高いストローク毎に算出した筆者内変動を示している．図 11 のように全員が自然筆跡の場合，それぞれにばらつきが存在した．それに対し図 12 では，筆者 A 本人の自然筆跡を除いて，他の合成偽筆の詐称者は筆者内変動が小さく，通常の自然偽筆に予測される筆者内変動の傾向とは逆の結果となった．これは尤度最大化基準に基づき抽出したパラメータをストロークに付加した結果，主な照合部位が付加パラメータ値の影響を共通に受けたためと考えられる．

４．５　まとめ

偽筆の定義を拡大し，人間の手による模倣を必要とせずに計算機上で生成する偽筆の可能性を検証した．その結果，テキスト独立型筆者照合システムの合成偽筆に対する脆弱性が

示され，閾値の変更のみでは合成偽筆を効率的に棄却することは困難である．そのため，合成偽筆を用いた詐称に対してロバストな筆者照合システムの検討を行う必要がある．

４．６　今後の課題

(1) テンプレート情報に対し透かしを埋め込む

今回提案した自然筆跡と合成偽筆の判別手法は，詐称者が筆者内変動を操作することにより判別不可の事態が予測される．そのため，詐称者による操作に左右されないなりすまし対策を施す必要がある．そこで，HMM を用いたテンプレート情報に対し透かしを埋め込むことで，漏洩し偽筆が生成されたとしても，透かし情報が偽筆データから抽出することができるような手法を提案する必要がある．また，オンライン情報が漏洩しHMM のテンプレート情報が推定される場合を想定し，オンライン情報に透かしを埋め込むことで，HMM に学習させた際にHMM に透かし情報が反映されるような手法についても検討を行う必要がある．

(2) 他の枠組みの筆者照合システムや偽筆合成システムについての検討

今回生成した合成偽筆が，現在製品化されている筆者照合システムに対しても適用可能か，また偽筆耐性がどの程度かについて検証する必要がある．

Ⅱ．顔照合の脆弱性に関する検討

１．背景と目的

顔情報を用いたバイオメトリック認証は抵抗感が少ない等の特長から，アクセスコントロールを実現するセキュリティ技術のひとつとして期待されており，実際に個人認証システムとして実用化され始めている．本稿では，バイオメトリクスの一つである顔情報に着目し，顔情報に基づく脆弱性のうち，特に，推定に関する脆弱性についてこれまで検討を加えた結果について報告する．

２．概要

２．１　推定に関する脆弱性と脅威

推定に関する脆弱性とは，テンプレートや照合結果から正当な登録者のバイオメトリックサンプルを推定できる脆弱性を示す．推定に関する脆弱性には，具体的な生体情報の推定方法として，

(i)　バイオメトリックシステムに保管されている正当な登録者のバイオメトリックテンプレートを何らかの方法で入手し，テンプレートに記述されているパラメータを元に顔情報を推定する方法

(ii)　照合結果として，スコア（類似度，距離など）を出力するようなバイオメトリックシステムに対し，入力顔画像を少しずつ変化させながら繰り返し照合し，スコアを改善させ，正しい生体情報に近い生体情報を作成する方法（ヒルクライミングアタック）

の 2 通りが挙げられる．すなわち，前者はバイオメトリックテンプレートから顔画像を推定する方式であるのに対して，後者はスコアを元に顔画像を推定する方式である．しかしテンプレートの持つ情報はアルゴリズムごとに異なり，顔認証アルゴリズムを提供するベンダー各社はテンプレート及びアルゴリズムの詳細な内容について公表していない．さらにテンプレートが持つ情報量は顔のそれよりもはるかに少ないため，例え前者のようにテンプレートを入手したとしても顔画像を推定することは困難であるといえる．一方，後者は照合結果であるスコアが得られる環境であるならば誰にでも推定攻撃を行うことができるという一面がある．実際バイオメトリック装置はセキュリティレベルの変化に柔軟に対応できるようにスコアが得られるようになっている物も存在しており，さらに BioAPI の標準化・運用が進むことにより，よりスコアの得やすい環境が整うと考えられる．
よって本研究では推定に関する脆弱性，そして，これを利用したヒルクライミングアタックを大きな脅威としてとらえ，検討を行った．

２．２　顔認証システムの仕様

顔認証システムの脆弱性を評価するにあたり，対象とする顔認証システムの仕様を明らかにする必要がある．本検討では顔特徴抽出アルゴリズムとして線形変換で評価のしやすい Eigenface 法[7]を用いる．

表 5 顔認証システムの諸元

学習顔画像	31 人×5 枚
累積寄与率（次元数）	94.9%（35）
スコア	クラス重心とのユークリッド距離（スコアが低いほど似ている）

２．３　顔推定システムの仕様

Adler の論文[6]を元に作成した．以下に概要を示す．

(1) 顔画像データベースを用意し，それに主成分分析を行うことで部分空間を作成する．
(2) 攻撃対象の ID を決める．
(3) 顔画像データベースの中から最もスコアの良い画像を探し，それを初期画像とする．
(4) 部分空間から任意の主成分（eigenface）を選択し，幾度かの試行を繰り返した後，最良の重みをつけた主成分を画像に加算する．
(5) スコアが改善された場合はその情報を保存し，悪くなった場合はやり直す．
(6) 目標のスコアになるまで(4)(5)を繰り返す．

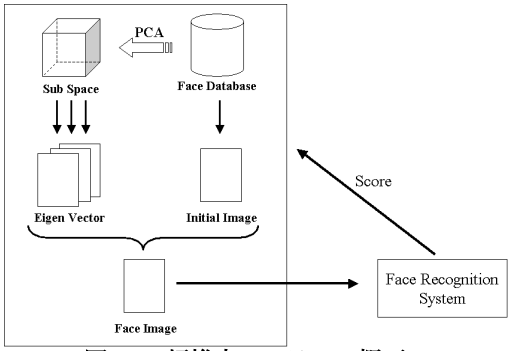


図 13　顔推定システムの概要

表 6 推定システムの諸元

学習顔画像	300 人×1 枚

累積寄与率（次元数）	99.0%（230）
目標スコア	300

３．ヒルクライミングアタック攻撃手法の効率化

Adler の手法を調査した結果，以下の効率化手法が考えられ，検討を行った．

- 初期画像生成手法
- 主成分の選択手法主成分の重み付け手法
- 領域分割による効率化手法

３．１　初期画像生成手法

３．１．１　提案手法

従来のヒルクライミングアタック手法では，攻撃対象である ID 決定後，その ID に対し攻撃データベースを総当りし，最良のスコアを取得した画像を初期画像としている．この従来手法に対し，本稿で用いた顔認証システムの顔特徴抽出アルゴリズムである Eigenface 法の特徴を利用し，各主成分の最適な重みを算出して重み付けすることで初期画像を生成する手法を提案する．

Eigenface 法では顔認証システムの顔データベースに対して，線形変換の主成分分析を用いて顔特徴を抽出（主成分を生成）し，テンプレート空間を作成する．また，照合の際には，入力画像に対して主成分分析を行い，顔特徴を抽出することでテンプレート空間に写像を行う．この照合の時，生成された入力画像の各主成分は，主成分分析が線形変換を行うため，テンプレートに対し線形な軸を持つと考えられる（図 14）．

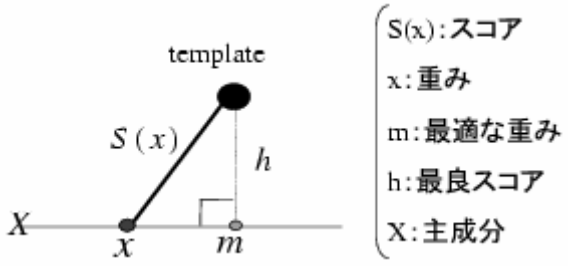


図 14　テンプレートに対する主成分の軸

図 14 より，スコア $S(x)$ は

$$S(x) = \sqrt{(x-m)^2 + h^2} \tag{3.1}$$

で表現される．式（3.1）より，各主成分のスコア $S(x)$ は，重み m の時に最もテンプレートに対する距離が短くなる（最良のスコアを取得する）ことが分かる．このことから，各主成分を，最良のスコアを取得する重み m を算出して重み付けし，画像に付加を行って初期画像を生成する．具体的な算出方法を以下に示す．主成分が 1 つ（1 次元）の場合（図 14），スコア $S(x)$ を算出する式は式（3.1）となる．式（3.1）において，算出する変数 m は

2 次の関数であることから，主成分軸上の 2 つの重み a，b を代入することで，算出できると考えられる．式（3.1）を変形し，重み a，b を代入した式は，

$$\begin{cases} S(a)^2 = h^2 + m^2 + a^2 - 2am \\ S(b)^2 = h^2 + m^2 + b^2 - 2bm \end{cases} \tag{3.2}$$

となり，この式（3.2）より，最良のスコアを取得する重み m は

$$m = \frac{S(a)^2 - S(b)^2 + b^2 - a^2}{2(b-a)} \tag{3.3}$$

で算出される．また，式（3.3）より，1 つの主成分（1 次元）の場合，m を算出するには $S(a)$，$S(b)$ の 2 つのスコアを必要とするため，2 回の認証回数が必要となる．

多次元の場合，各主成分の最適な重み m を算出することとなるため，各主成分軸上の 2 つの重み a，b を代入する．n 次元の場合の各主成分の最適な重み m は

$$\begin{cases} m_1 = \dfrac{S(a_1,a_2,\ldots,a_i,\ldots,a_n)^2 - S(b_1,a_2,\ldots,a_i,\ldots,a_n)^2 + b^2 - a^2}{2(b-a)} \\ m_2 = \dfrac{S(a_1,a_2,\ldots,a_i,\ldots,a_n)^2 - S(a_1,b_2,\ldots,a_i,\ldots,a_n)^2 + b^2 - a^2}{2(b-a)} \\ \mathrm{M} \\ m_i = \dfrac{S(a_1,a_2,\ldots,a_i,\ldots,a_n)^2 - S(a_1,a_2,\ldots,b_i,\ldots,a_n)^2 + b^2 - a^2}{2(b-a)} \\ \mathrm{M} \\ m_n = \dfrac{S(a_1,a_2,\ldots,a_i,\ldots,a_n)^2 - S(a_1,a_2,\ldots,a_i,\ldots,b_n)^2 + b^2 - a^2}{2(b-a)} \end{cases} \tag{3.4}$$

で算出される．また，式（3.4）より，n 次元の場合，各主成分の最適な重み m を算出するには $n+1$ 個のスコアが必要となり，伴って認証回数も $n+1$ 回必要となる．

この手法で考えられるメリットとして，従来手法では攻撃データベース内全ての画像枚数の認証回数が必要となるが，提案手法では入力画像の次元数の認証回数で初期画像を生成でき，また各主成分をテンプレートに対する最適な重みで重み付けしているため，従来手法よりも良いスコアからヒルクライミングアタックを開始できる．

３．１．２　結果・考察

従来手法と提案手法の初期画像におけるスコアの比較結果を以下に示す．

表 7　初期画像生成手法:スコア比較

ID.NO	従来手法	提案手法
ID.1	1143	686
ID.2	972	615
ID.3	1345	975
ID.4	1457	839
ID.5	786	513

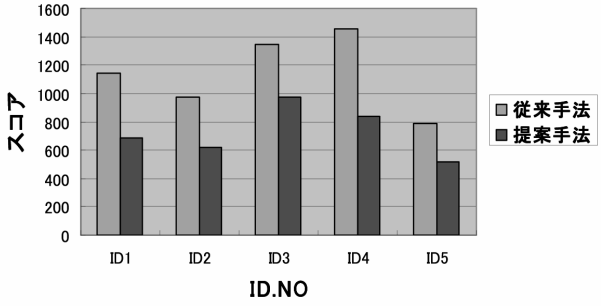


図 15　初期画像生成手法：スコア比較

表 7，図 15 より，どの ID に対しても初期画像において取得するスコアが従来手法より 60～70%程度に減少され，効率化された．しかし，各主成分をテンプレートに対して最適な重みで重み付けしているにも関わらず，スコアは約 500～1000 と目標スコア（300）に対し必ずしも近くない距離となった．その理由として，最適な重み m を算出する際に，照合時における主成分軸のテンプレート空間への写像を考慮に入れていないことが考えられる．各主成分は元空間ではテンプレートに対して様々な方向に直交しながら線形な軸を持っている．そして照合時にはそれらの軸はテンプレート空間に写像される．その際，各主成分のテンプレートに対する軸の方向は変化し，軸とテンプレートとの距離も変化すると考えられる．よって，最適な重み m を算出するにはテンプレート空間への写像，そしてその際に生じる各主成分軸の相関関係も考慮せねばならない．

３．２　主成分の選択手法

３．２．１　提案手法

従来のヒルクライミングアタック手法では画像に加算する主成分を選択する際，その主成分をランダムで選択している．この従来手法に対し，各主成分の様々な重みにおけるスコア推移に着目し，スコアに対して影響力の大きな主成分から選択する手法を提案する．

式（3.1）より，各主成分を様々な重みで画像に加算した場合のスコア推移は，重み 0 を交点とする 2 次曲線となると分かる．また，各主成分において，微小な重みの変化では劇的な画像の変化は期待されないため，その場合にスコアはほぼ直線として推移すると考えられる．この時，その各主成分の直線的な推移は，テンプレートに対する主成分軸の距離が異なるため，様々な傾き（変化量）を持っていると考えられ，この傾きの大きな主成分ほどスコアに対し大きな影響力を持つと思われる．よって，各主成分に対し，初期画像におけるスコアであり，各主成分共に等しくなる重み 0 と，微小に異なる重み x でのスコアの推移から傾きを算出し，その傾きの大きな主成分から選択する．

この手法のメリットとして，スコアに対し影響力のある主成分から選択しているため，目標スコア収束までの主成分選択回数を減少させることができると考えられるが，一方で，各主成分のスコア推移の傾きを算出するため，n 次元の入力画像では，n 回の認証回数が必要となるデメリットもある．

３．２．２　結果

従来手法と提案手法の主成分選択回数における比較結果を以下に示す．

表 8 主成分選択手法:主成分選択回数比較

ID.NO	従来手法(回)	提案手法(回)
ID.1	224	156
ID.2	187	162
ID.3	283	231
ID.4	262	173
ID.5	156	102

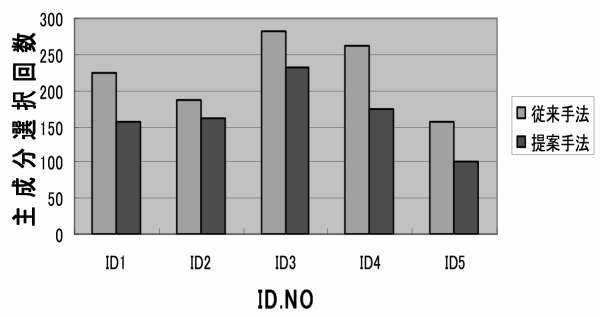


図 16 主成分選択手法：主成分選択回数比較

表 8, 図 16 より，どの ID.NO に対しても目標スコア収束までの主成分選択回数は 70~80%程度に減少された。

３．３ 主成分の重み付け手法

３．３．１ 提案手法

従来のヒルクライミングアタック手法では，選択した主成分をランダムで決定した重みで重み付けする試行を複数回行い，最終的に最良のスコアを取得した重みで重み付けする（図 17）．この従来手法に対し，節 3.2 と同様，各主成分のスコア推移に着目し，その推移を $S(x)=ax^2+bx+c$ という形の 2 次関数に近似して，主成分の極となる重みを算出して重み付けする手法を提案する．

式（3.1）より，各主成分を様々な重みで画像に加算した場合のスコア推移は 2 次曲線となると分かる．このことから，各主成分のスコア推移を $S(x)=ax^2+bx+c$ の 2 次関数に近似し，選択した主成分の 3 つの重みにおけるスコアから，その主成分の極となる重みを算出し，重み付けする（図 18）．

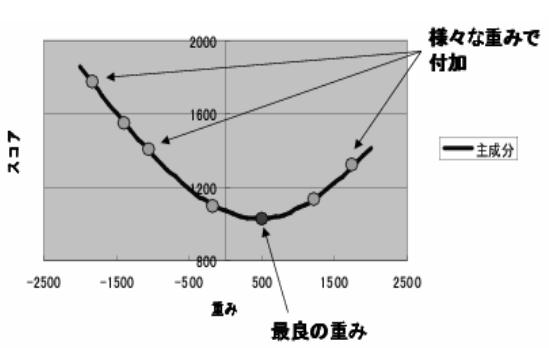


図 17 従来手法の重み付け手法

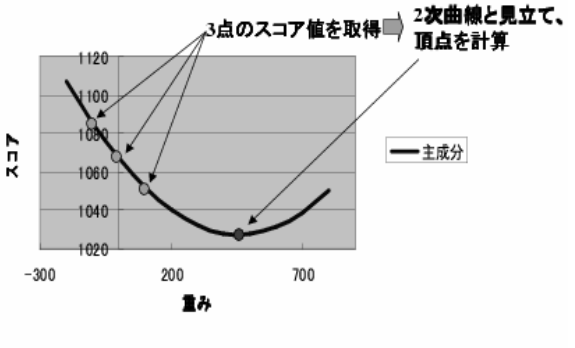


図 18 提案手法の重み付け手法

この手法のメリットとして，1 つの選択した主成分に付き，3 回の認証回数で済み，また算出した重みで重み付けすることから効率的にスコアを改良させ，認証回数の減少につながると考えられる．

従来手法における最も効率的な重みの付加試行回数を調査した結果，付加試行が 2 回の時

に最も早く目標スコアまで収束したため，そのデータを提案手法に対する比較対象とした．

３．３．２　結果・考察

従来手法と提案手法の認証回数における比較結果を以下に示す．

表 9　主成分の重み付け手法:認証回数比較

ID.NO	従来手法(回)	提案手法(回)
ID.1	748	563
ID.2	674	576
ID.3	866	801
ID.4	824	733
ID.5	612	498

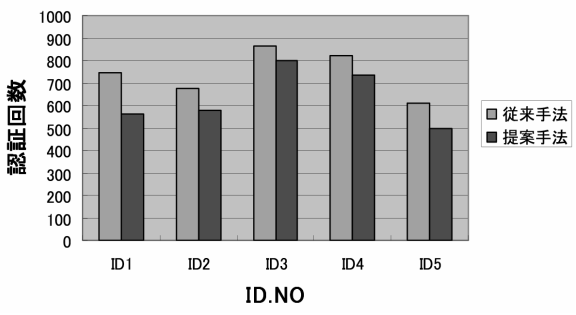


図 19　主成分の重み付け手法：認証回数比較

表 9 と図 19 より，どの ID.NO に対しても目標スコア収束までの認証回数は 75~90%程度に減少した．選択した主成分に対する重みの付加試行回数は従来手法が 2 回であるのに対し，提案手法では 3 回とその回数は増えている．しかしながら，従来手法では主成分をランダムで決定した値で重み付けしているため，必ずしもスコアが改善されるとは限らず，選択が無駄となる主成分が多いと考えられる．それに対し，提案手法では選択した主成分のスコアが最良となると考えられる値を算出して重み付けし画像に加算しているため，効率よくスコアを改善できている．結果として，目標スコア収束までの認証回数は提案手法の方が少なくなったと考えられる．

３．４　領域分割による効率化手法

３．４．１提案手法

Adler の手法では頭部領域全体を一つの画像として処理を行っている．しかし顔の中には目や口のように認証に大きく影響すると考えられる領域と，肌のように影響が少ないと考えられる領域が存在する．そこで画像を複数領域に分け，認証に大きく影響する領域の処理回数を増やすことで，より効率的にヒルクライミングアタックができるのではないかと考えられる．

これを確認するために，本実験では画像を様々な大きさの格子状に分割し，乗算・加算回数を計算量として求めている．また領域の重み付けは 3.4.1.1 で提案する手法を用いている．

3.4.1.1 重み付け手法

重み付けを考えるにあたり，まず分割した領域を処理することでどのような変化が起こっているのかを考察する．Adler の手法では初期画像に主成分を加算することで画像を処理している．よって，分割された領域を処理することは，分割した主成分を加算することを意味している．言い換えれば，加工対象とする領域以外を 0 にマスクした主成分を加算することになる．ここで元の第 i 主成分を EF_i とした時，特定の領域 m に合わせるためにマスク

した主成分を MEF_{mi} と呼ぶことにする．

初期画像にマスク主成分 MEF_{mi} を加算した場合，特徴空間上の 1 点である初期画像は MEF_{mi} のベクトル方向を移動することになる．従って，移動方向に攻撃対象のテンプレートに対応する領域が存在するならば，移動によってスコアは変化することになる．この様子を図 20 に示す．全ての MEF_{mi} の長さを 1 になるように正規化した場合，この移動による変化値が大きいほど，その MEF_{mi} のベクトル方向はテンプレート対応領域の方向を向いており，スコア改善を期待できるマスク主成分ということになる．以上の考察から，式 (3.5) で求まる領域有効値の割合で実際に処理する領域を決定し，重み付けを行った．

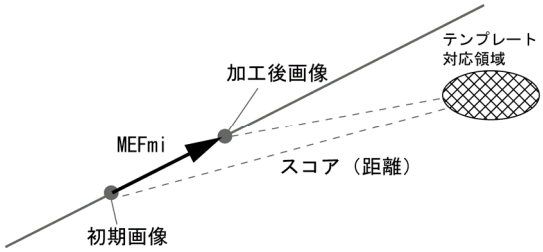


図 20　MEF_{mi} を加算した様子

$$Score_m = \frac{\sum_i^{EF_{\max}} \left| Score(IM_{init}) - Score(IM_{init} + MEF_{mi}) \right|}{EF_{\max}} \quad (3.5)$$

３．４．２　実験・考察

6 種類の分割方法に対して提案した重み付け手法を適用し，それぞれの計算量（乗算回数）を比較した．実験結果を表 10 と図 21 に示す．

表 10 各手法の演算回数

手法	平均認証回数	平均乗算回数
分割なし	766.9	2.163×10^8
12 分割	811.5	2.066×10^8
48 分割	795.4	2.010×10^8
300 分割	684.1	1.725×10^8
1200 分割	619.9	1.563×10^8
7500 分割	861.0	2.170×10^8

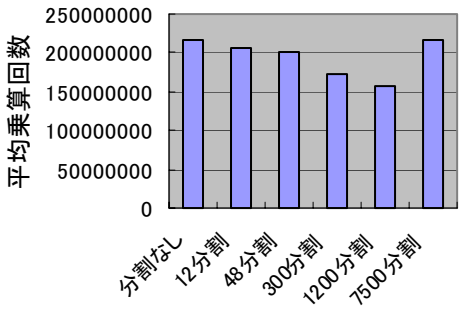


図 21　各分割方法の計算量

実験の結果から計算量を最大 72%まで抑えることに成功し，領域分割により効率化が可能であることが分かった．

４．ヒルクライミングアタックの対策

ヒルクライミングアタックの対策手法として，以下の 2 つを検討した

・ 認証スコアから漏洩する情報を減らす手法
・ ヒルクライミングアタックの特徴を利用する手法

４．１　認証スコアから漏洩する情報を減らす手法

４．１．１　提案手法

認証器から出力されるスコアにはテンプレートの情報を推定できる情報が含まれている．この脆弱性に対し，出力される認証スコアを量子化することで漏洩する情報を減らし，攻撃を防ぐ対策手法が提案されている[8]．しかし量子化による対策手法に対しては，既に Adler により改良版ヒルクライミングアタック[9]が提案されている．そこで本節ではスコア曲線を見直し，Adler の攻撃手法に対しても有効な防御手法を提案する．

従来の量子化手法が

$$y = z(x) \tag{4.1}$$

と表されるとすると，本提案手法は次のように表される．

$$y = 2z(x) - x \tag{4.2}$$

これを図 22 に示す．

４．１．２　実験・考察

量子化手法，提案手法に対して，改良版ヒルクライミングアタックを行った．実験結果を図 23 に示す．この結果から量子化手法は収束してしまっているのに対して，提案手法では収束が抑制されており，提案手法は改良版ヒルクライミングアタック手法に対しても耐性があることがわかった．

ただし今回提案したスコア曲線は量子化手法よりは複雑であるものの，やはり単純な曲線である．そのためテンプレートを推定できる情報が残っている可能性があり，攻撃の可能性を否定できない．どのような部分に脆弱性が存在するのかを調査し，テンプレートを推定できる情報が残らないようなスコア曲線を提案する必要があるといえる．

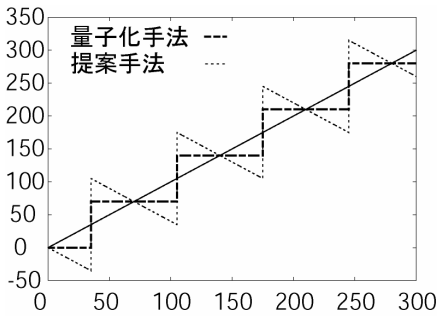


図 22　スコア曲線

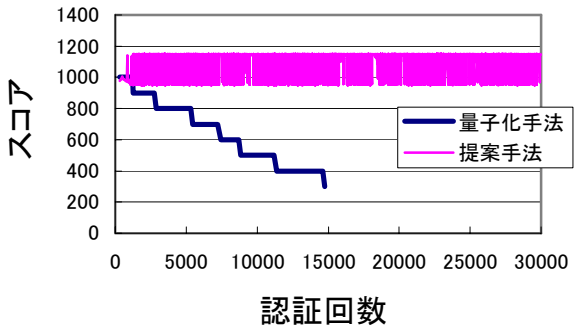


図 23　改良版ヒルクライミングアタックに対する対策手法の効果

４．２　ヒルクライミングアタックの特徴を利用する手法

４．２．１　提案手法

推定に関する脆弱性を利用した攻撃手法はヒルクライミングアタックである．このヒルクライミングアタックで作成した画像には不自然さが存在するという特徴がある．これはヒルクライミングアタックにおける画像の評価の仕方が，実際の見た目でなく，認証スコアであるためである．そこでこの不自然さを利用することで，ヒルクライミングアタックで作成された画像を見分けることができるのではないかと考えられた．

そのためには一般的な不自然さ（偽造）を定義しなくてはならないが，画像の不自然さは顔認証アルゴリズムごとに異なり，一般的な不自然さを定義するのは難しい．そこで本稿では逆転の発想から，先に不自然な画像を定義し，ヒルクライミングアタックを行うと，そのような画像が推定されるように誘導することを考えた．

通常スコア曲面は図 24 のようにテンプレートに近づくにつれてスコアが良くなっている．そのためヒルクライミングアタックを実行するとテンプレートに近づくように収束する．そこでスコア曲面が図 25 のようになるようにし，ヒルクライミングアタックを行った際にはテンプレートから遠ざかり，偽造領域に収束するようにさせた．なお本稿では不自然さを見分けるパラメータとして高次元の主成分の値を用いている．

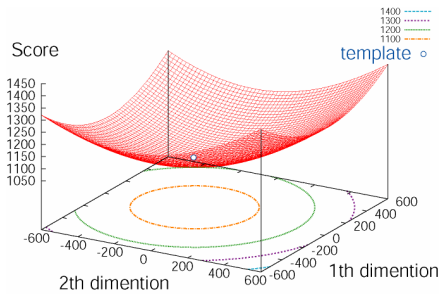


図 24　通常のスコア曲面

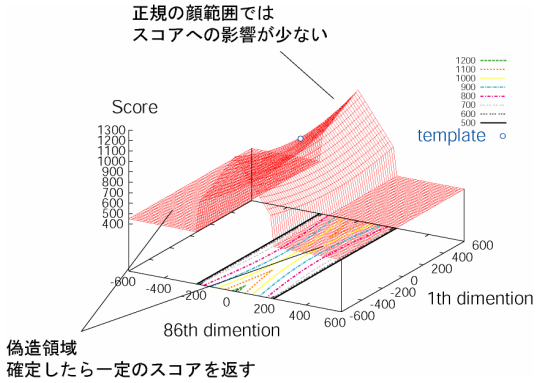


図 25　提案するスコア曲面

４．２．２　実験・考察

提案した対策手法を施した認証器に対して，Adler の手法でヒルクライミングアタックを行った．実験は 31 名の被験者に対して各 10 回，計 310 回行った．

この実験の結果，提案手法は 310 回中 290 回に対して有効に働き，ヒルクライミングアタックを防ぐことに成功した．またこの対策手法による認証精度への影響は少ないことを確認している．成功時の収束の例を図 26 に，失敗時の例を図 27 に示す．

ただし本提案手法は同一人物に対しても成功する時としない時があった．これは仮にヒルクライミングアタックを防ぐことができても，何回もトライされれば成功されてしまう可能性があることを意味している．さらに今回の手法では高次元の主成分の値を不自然さを見分けるパラメータとして使っているため，攻撃者が高次元の主成分を使わないようにフ

ィルタすることで攻撃が可能になってしまう．これに対し低次元の主成分を不自然さを見分けるパラメータとして使用する方法もあるが，低次元は顔認証でも使われている可能性の高いパラメータなため，不自然さを見分けるためにそのパラメータを変化させることは，認証精度に大きく影響してしまう可能性がある．

今回の手法を成功させるためには，確実に推定画像を偽造領域に収束させることと，認証にあまり用いておらず，攻撃者が操作できないパラメータを探し出し，それを不自然さを見分けるパラメータとして使用することが必要だと考えられる．

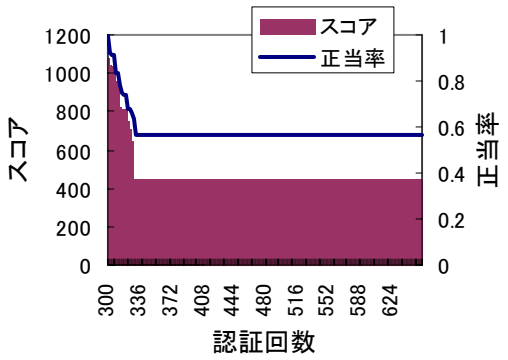


図 26　成功例

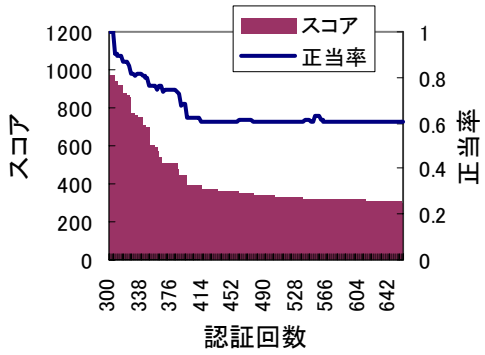


図 27　失敗例

５．まとめ

本稿では顔認証に関する脆弱性のうち，推定の脆弱性を最もリスクの高い脆弱性と考え，検討を行った．検討内容は「脆弱性の程度の把握」と「脅威に対する対策手法の検討」に分けられる．

脆弱性の程度の把握では　脆弱性の程度を調べるにあたり，現状の推定の脆弱性に対する攻撃手法であるヒルクライミングアタックの効率化を検討し，その攻撃可能性を調査した．その結果，効率化は可能であることを確認し，今後脆弱性の程度を評価する際にはこのことを考慮に入れるべきであると考えられた．

脅威に対する対策手法の検討では 2 つのコンセプトから対策手法を検討し，それぞれの手法の可能性を示すことができた．

６．今後の課題

Eigenface 法を評価対象とし，ヒルクライミングアタックの効率化を検討した結果，提案した各手法にて効率化を達成することができた．これは対象とした Eigenface 法が線形変換を用いているにも関わらず距離をそのままスコアとして出力しているため，スコアから生体情報が推定しやすかったと考えられる．このように入力データと学習データの距離関係がそのままスコアとして出力される認証システムは脆弱であると考えられ，スコア量子化や偽造認識などの対策手法の検討が必要と考えられる．

また本稿では触れなかったが，提案した効率化手法を顔認証製品に適用したところ，期待されたほどの効率化は得られなかった．これは今回検討対象とした Eigenface 法が線形変換であるのに対し，この顔認証製品は非線形の特性を有する結果であったためと考えられる．

しかし，このような非線形変換の認証器に対しても，入力画像の変動とスコアの関係の分析を進めることにより効率化を行う手法が存在すると考えられ、こうした特性を有する認証器も含めた検討を行う必要がある．

Ⅲ．話者照合の脆弱性に関する検討

１．目的

本稿においては，携帯電話やＩＰネットワーク上のディジタル音声通信において音声符号化方式との親和性を考慮した話者照合方式(以下，CELP 話者照合方式)を提案し，その提案システムを用いて脆弱性の検討を行う．特に，推定に関する脆弱性を扱い，これまで検討を加えた結果を報告する．

２．CELP 話者照合方式

CELP 話者照合方式[10,11]では，携帯電話の音声通信や VoIP(Voice over IP)で利用されている CELP(Code Excited Linear Prediction:符号励振線形予測)符号化方式)[12]により符号化された音声情報を使用して話者照合を行う．CELP 話者照合方式は，以下のような特徴を持つ．

１．　符号化された音声情報のみを用いて話者照合を行うため，端末側，ネットワーク側(センタ側)のいずれにおいても話者照合を行うことができる．

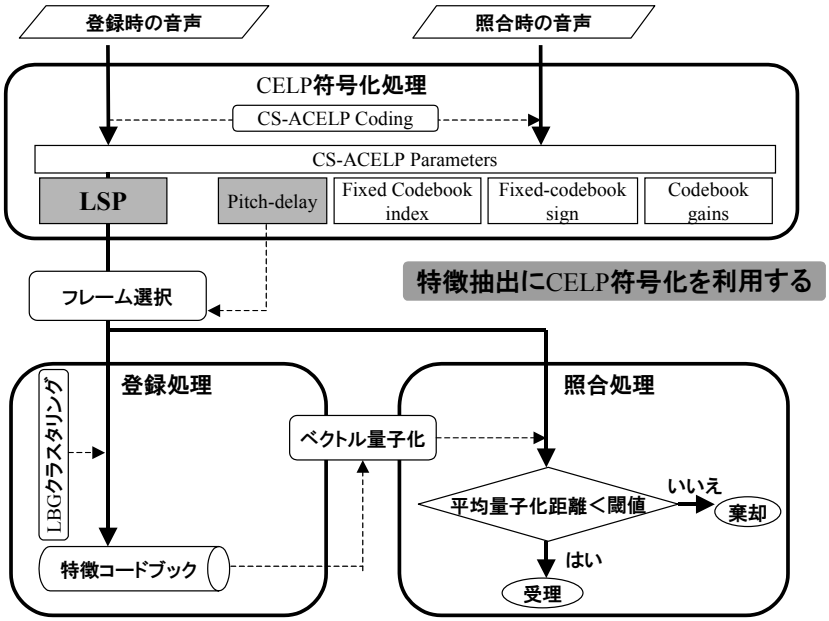
２．　移動通信システムに適用した場合，端末に搭載されている音声符号化機能を利用できるため，話者照合のために必要となる機能の追加を抑えることが可能であり，重量やサイズに制限のある端末側での認証に有利である．

３．　調音動作を反映するパラメータを使用することにより，話者照合の発話内容に依存しないテキスト独立型の話者照合が可能となる．

提案する CELP 話者照合方式の手順について説明する．提案方式は，CELP 符号化プロセスと話者照合のための登録・照合プロセスからなる．まず，CELP 符号化プロセスでは，入力音声を CELP 符号化してパラメータを抽出する．この符号化パラメータの１つである LSP が話者照合に利用される．登録プロセスでは，登録用の LSP をクラスタリングによって，個人性を抽出し，コードブックを作成する．このユーザの個人性を表すコードブックを特徴コードブックと呼ぶ．この際，雑音に強い話者照合を実現するために，環境雑音に対して安定な特徴を示すフレームを選択する処理を行う．環境雑音に対して LSP の変化が小さいフレームを選択するにあたり，CS-ACELP[13]のための無音検出アルゴリズムと CS-ACELP パラメータの一つであるピッチ(Pitch delay)を利用することで，雑音に強いフ

レームを選択している．照合プロセスでは，照合用の LSP を特徴コードブックによってベクトル量子化し，量子化歪みに基づいて受理・棄却の判定を行う．

提案方式では，CELP 話者照合方式の一種であり，ITU/T で G.729 として標準化されている CS-ACELP(Conjugate Structure Algebraic CELP:共役構造代数 CELP)を使用している．



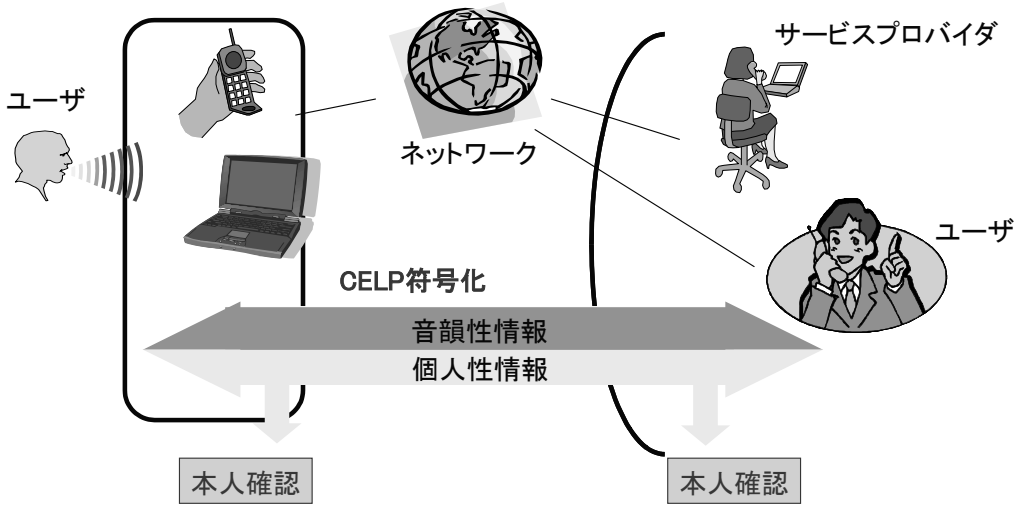
提案方式の概要を図１に示す．

図１：CELP 話者照合方式の概要

３．CELP 話者照合方式を用いたシステム

CELP 話者照合システムの構成例を図２に示す．このシステムは，電子商取引などを提供するサーバに CELP 話者照合方式を実装し，照合中の音声を用いてサービス利用者の真正性を確認するものである．クライントには携帯電話やＰＣ等が用いられるが，CELP 符号化処理は携帯電話や VoIP クライアントに標準で実装されているため，照合のために特に追加する機能を必要としない．

クライアントで取得した音声は，CELP 符号化されてネットワークに送信される．サーバでは受信した音声の音韻性情報を用いて，取引の内容を確認すると同時に CELP 符号化された情報から個人性情報を抽出し，利用者が登録された本人であるか否かが確認できる．



例えば携帯電話の通話中の認証を行い，携帯電話の不正利用を防止したり、ネットワーク会議などのアプリケーションで，登録されたユーザのみが利用する等の応用が考えられる．

図２：CELP 話者照合方式

４．推定に関する脆弱性

４．１　バイオメトリクス装置における推定の脆弱性となりすまし

バイオメトリクス装置に保管された正当な利用者のテンプレートもしくは照合結果から，利用者の生体情報を推定できる脆弱性を「推定」と呼ぶ．攻撃者は，テンプレートを取得・解析できれば，利用者の生体情報を推定し，利用者になりすますことが可能になる．また，テンプレートの解析が不可能でも，テンプレートと入力生体情報の間の照合結果として，スコア(類似度)を得ることができる環境を持っていれば，繰り返し生体情報を入力することで，スコアの高くなる生体情報を推定することができる．この攻撃をヒルクライミングアタック[14]と呼ぶ．

４．２　話者照合システムに対するなりすまし

話者照合システムにおける推定の脆弱性を利用したなりすましの方法は，

A)　テンプレートを解析して，利用者の音声を推定する

B)　ヒルクライミングアタックを用いて，利用者のスコアに近づくような音声を推定するの 2 通りを考えることができる．A のようにテンプレートを解析するためには，対象とする話者照合システムが使用する照合アルゴリズムやテンプレートのデータ形式に関する知識が必要となる．一方，B の方法では，アルゴリズムやテンプレートについて未知であっても，攻撃可能であるが，計算コストは一般に A より大きいと考えられる．

話者照合システムは，「テキスト依存型」・「テキスト独立型」・「テキスト提示型」の３つのタイプに分けることができる．各タイプ別になりすましの手法は異なるが，いずれの場合も推定に利用する音声は音声合成もしくは声質変換の技術を応用して作成することが可能であると考えられる．テキスト依存型に対するなりすましでは，固定されたテキストに対する利用者の音響的特徴をテンプレートもしくはスコアから推定して音声を合成する．テキスト独立型に対するなりすましは，テキストによらない(様々なテキストに対応する)音響的特徴を推定し，任意のテキストを用いて音声を合成する．テキスト提示型に対しては，テキスト独立型と同様にテキストによらない音響的特徴を推定し，照合時にシステムから提示されたテキストを用いて音声を合成する．

４．３　CELP 話者照合システムに対するなりすまし評価実験

本稿では，テキスト独立型の話者照合方式である CELP 話者照合方式を対象として，なりすましを行う手法について検討した結果を報告する．

今回は，テンプレートや対象となる話者照合アルゴリズムについて既知であることを前提として，テンプレートを解析して利用者の音声を推定する方法について検討した．推定に利用する音声は，攻撃者が適当に用意した音声から声質変換の手法を応用して利用者の音声に変換する方法をとった．

５．実験結果

なりすましの方法，及び評価実験の結果を以下に示す．

５．１　なりすましの方法

入力音声を線形予測分析し，調音情報である LSP と音源情報に対応する予測残差とに分解する．LSP は，盗まれたテンプレートによってベクトル量子化され，利用者の特徴に置き換えられる．置き換えられた LSP と予測残差を用いて音声合成[15]することで，偽声を生成する．なりすましのための偽声生成手順を図３に示す．

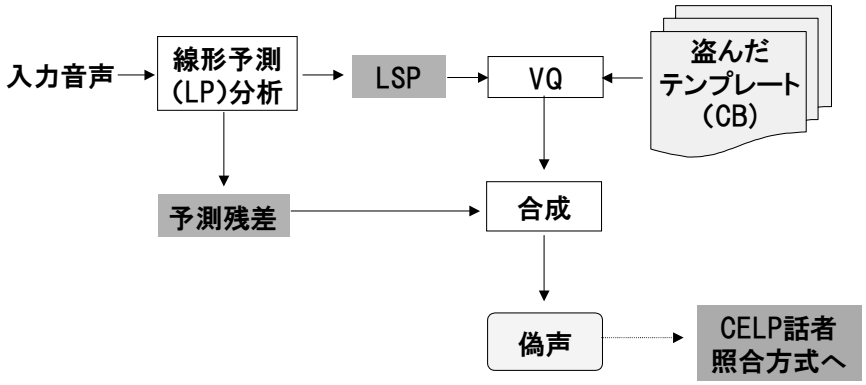
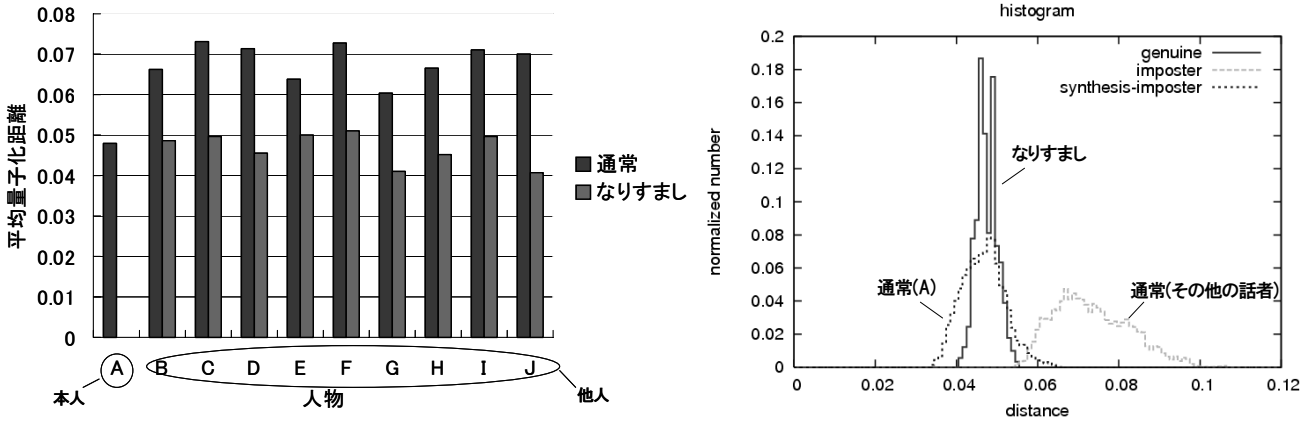


図３　なりすましのための偽声生成手順

５．２　評価実験



(a)話者ごとの平均量子化距離

(b)距離の分布

図４　ある女性話者 A に対するなりすまし時の照合結果

図４(a)に CELP 話者照合方式におけるベクトル量子化の際の平均量子化距離を示した．B～J のすべての話者が，A になりすました場合の平均量子化距離が A 自身の音声と同等またはそれ以下となった． 図４(b)に量子化距離の分布を示す．通常の場合(なりすましを行わない場合)は、A とその他の話者との分布がわかれているが，なりすましを行うと A の分布に重なる．これらの結果から、本手法によってなりすましの危険性が十分存在することが分かる．

６．まとめ

本稿では，CELP 符号化された情報から個人性を抽出して話者照合を行う方式(CELP 話者照合方式)を利用し，その提案システムの脆弱性について検討した．特に「推定」に関する脆弱性を扱い，テンプレートの内容及び照合アルゴリズムが既知であるとしてなりすましの可能性を検証した．その結果、なりすましができる危険性の存在が確認できた．

７．今後の課題

・ヒルクライミングアタック

テンプレート内容及び照合アルゴリズムが未知であっても音声の推定を可能とするヒルクライミングアタックの手法について検討し，なりすましの危険性について検証する．また，それを踏まえた上でヒルクライミングアタックを防ぐ手段を提案する．

・なりすまし対策

本稿では，合成音声によるなりすましの危険性を確認した．しかし，再現の難しいパラメータの利用，テンプレート内へのダミーデータの混入などの手法を用いることで，なり

すましされた音声であることを検知して棄却できる可能性がある．今後，具体的ななりすまし防止アルゴリズムについて検討を行う．

・その他の脆弱性

話者照合システムに考えられる脆弱性として、「推定」の他にも「不定データ」「複製」「秘匿困難」「変化」などが挙げられる．今後は，「推定」の脆弱性にとどまらず，システムの脆弱性全般について検討する予定である．

参考文献

[1]吉村ミツ，吉村功，``筆者認識研究の現段階と今後の動向”，信学技報，PRMU96-48, pp.81-90，1996.

[2]社団法人日本自動認識システム協会,``生体情報による個人識別技術（バイオメトリクス）を利用した社会基盤構築に関する標準化”，2004.

[3]Mitsu Yoshimura, Yutaka Kato, Shin'ichi Matsuda, Isao Yoshimura, ``On-line Signature Verification Incorporating the Direction of Pen Movement”, IEICE Trans. E74, 7, pp.2083-2092, 1991.

[4] 小林隆夫, 徳田恵一, ``コーパスベース音声合成技術の動向[IV]－HMM 音声合成方式－”, 電子情報通信学会誌 Vol.87，No4，2004.

[5] 中川研究室. “標準手書きデータベース HANDS-nakayoshi-d”．東京農工大学.

[6] Andy Adler, “Sample Images can be Independently Restored from Face Recognition Templates",CCECE 2003- CCGEI 2003,Montreal,May/mai 2003[7]Matthew A.Turk and Alex P.Pentland，”Face Recognition Using Eigenfaces”，IEEE 1991

[8]Colin Soutar, ”Biometric system performance and security”, Mytec Technologies Inc., http://www.bioscrypt.com/assets/bio_paper.pdf

[9]Andy Adler, ”IMAGES CAN BE REGENERATED FROM QUANTIZED BIOMETRIC MATCH SCORE DATA”，IEEE 2004

[10]T.Mogaki and N.Komatsu : ”Text-indicated speaker verification method using PSI-CELP parameters”, Proc.of SPIE, Vol.3657, pp.184-193,1999.

[11]山崎恭, 近藤維資, 小松尚久：”CELP パラメータを用いた話者照合方式”, 画像電子学会誌, Vol.32, No5, pp.629-634, 2003.

[12]ITU-T : “Coding of speech at 8 kbit/s using conjugate structure algebraic code- xcited linear prediction(CS-ACELP)”, ITU-T Recommendation G.729, 1996.

[13]ITU-T : ”Coding of speech at 8 kbit/s using conjugate structure algebraic code excited linear prediction(CS-ACELP) Annex B : A silence compression scheme for G.729 optimized for terminals conforming to Recommendation V.70”, ITU-T Recommendation G.729(Annex B), 1996.

[14]Colin Soutar. “Biometric system performance and security”. Sep. 1999.
[15]益子貴史, 徳田恵一, 小林隆夫：”話者照合システムに対する合成音声による詐称”, 信学論Ｄ－Ⅱ, Vol.J83. No.11, 2000.