画像記憶のスキーマを利用したユーザ認証システム

原田篤史[†] 漁田武雄[††]
水野忠則[††] 西垣正勝[††]

現行のユーザ認証方式は，汎用性と利便性の高さから文字をベースとしたパスワード方式が主流となっているが，人間にとって長くランダムな文字列を記憶することは容易ではなく，新たなパスワードを設定するたびにユーザは記憶に関する大きな負担を要求される．そのため，人間の画像認識能力の高さを利用して認証情報の記憶の負荷を低減させる画像認証方式が注目されている．しかし，多くの画像認証方式は毎回の認証時にパスワード画像がディスプレイ上に表示されるため，認証時の覗き見攻撃に対して脆弱であった．そこで本論文では，有意味なオリジナル画像に対してモザイク化をはじめとする不鮮明化処理を施すことで一見して無意味な画像を作成し，それをパスワード画像として用いることで覗き見攻撃への耐性を有する画像認証方式を提案する．正規ユーザは，パスワード画像登録時に不鮮明化処理を行う前のオリジナル画像を見ることができるため，不鮮明化されたパスワード画像をオリジナル画像の持つ意味と結び付けて記憶することができる．そのため，認証試行時には画面上に表示される自身の不鮮明なパスワード画像を容易に再認識することが可能となり，記憶の負荷は小さくなる．一方，オリジナル画像を見ることのできない他のユーザが当該ユーザの認証行為を覗き見たとしても，意味を認識できない不鮮明なパスワード画像を記憶にとどめておくことは困難である．また，他者が正規ユーザからパスワード画像の意味を言葉で教えられたとしても，不鮮明なパスワード画像を教えられた意味どおりに認識することは困難であるため，本方式は正規ユーザからのパスワード漏洩に対してもあるレベルの耐性を有する．

A User Authentication System Using Schema of Visual Memory

Atsushi Harada,[†] Takeo Isarida,[††] Tadanori Mizuno[††]
and Masakatsu Nishigaki[††]

Although password-based systems are now widely used in all kinds of authentication, they have a problem that humans have a limitation to remember secure passwords (long and random strings). Thus, image-based user authentication systems using “pass-images” instead of passwords have been studied for reducing the burden of memorizing passwords. However, on many image-based systems, it is needed to present a user’s pass-image on their display at each authentication trial, so they can be vulnerable against an observing attack. In this paper, we propose a user authentication system using “unclear images” as pass-images, in which only the legitimate users are allowed to see the original (picture) images corresponding to their unclear pass-images in the enroll phase. The legitimate users can easily remember their unclear pass-images using the original images as clues, while illegal users without the clues have difficulties to find out and remember other user’s unclear pass-images. In addition, it is expected to be difficult for even a legitimate user to leak his/her unclear pass-image precisely to anyone with words via e-mail or telephone.

1. はじめに

現在のユーザ認証は汎用性と利便性の高さから，パスワードや暗証番号のような，文字や記号の記憶を利用したユーザ認証方式が主流となっている．この場合，できるだけ長く，ランダムな文字や数字を用いるほど安全なパスワードとなるが，本来，人間は長い文字列や記号列を正確に記憶することが得意ではない．そのため，長くランダムなパスワードを設定することを躊躇したり，定期的に新たなパスワードを覚えなおすことに苦痛を感じたりすることがある．実際，多くのユーザがパスワードに名前や誕生日などを用いたり，パスワードを紙に書きとどめておいたり，様々な環境で同一のパスワードを使いまわしたりする傾向があることが知られている．そのようなパスワードは辞書攻

† 静岡大学大学院理工学研究科
Graduate School of Science and Engineering, Shizuoka University
†† 静岡大学情報学部
Faculty of Informatics, Shizuoka University

撃や身近な人物によるソーシャルエンジニアリングなどに対して脆弱であり，システムのセキュリティレベルを低下させる原因となっている[1),16)]．

この問題の解決に向け，人間が得意とする画像記憶を用いて認証情報の記憶負荷を軽減する画像認証方式が多数提案され，注目されている[2)~9)]．特に，認証時に提示される複数の画像中に含まれる自身のパスワード画像を選択することで認証が行われる方式[5)~9)]では，画像記憶の利用に加え，再認によってパスワードの想起が容易になっており，ユーザに要求される記憶負荷が大きく抑えられている．

しかし通常の画像認証方式には，「毎回の認証時にパスワード画像がディスプレイ上に表示されるために，認証行為を覗き見されてしまうと，パスワード画像が漏洩する危険性がある」という画像による認証を行うがゆえの欠点が存在する．画像の使用は正規ユーザの記憶負荷を低減させるが，それは同時に，攻撃者にとっても覗き見た他人のパスワード画像を容易に記憶可能であるということであり，画像認証方式における覗き見攻撃（ショルダーハッキング）の脅威は大きい．

他者による覗き見を防止する対策として，商用に広く販売されている狭可視角化フィルタ[18)]をディスプレイに装備することが考えられる．これは簡単で効果的な方法であるが，横からの覗き見は防止できても，後ろからの覗き見までを防ぐことができない．文献 19) は後ろからの覗き見にも効果がある方式であるが，認証情報を復号するための特殊なスライドシートを所持する必要がある．また，専用メガネをかけないとディスプレイ上の表示が見えない PPT（Picture Protect Technology）ディスプレイ[11)]においては，同じ PPT 専用メガネを所有する人物によって画面表示を見られてしまうという問題が指摘されている[12)]．覗き見を簡単かつ確実に阻止することは物理的に困難であるため，記憶の負荷を低減しつつ，覗き見攻撃にも耐性を有する画像認証方式が望まれる．

その実現に向けて，本論文では，正規ユーザ本人には記憶が容易であり，他者にとっては記憶が困難となるような画像をパスワード画像として使用する画像認証方式を提案する．すなわち，記憶の負荷を低減しつつ，覗き見攻撃にも耐性を有する素材を使用しての画像認証方式の実現を目指す．具体的には，有意味なオリジナル画像に対してモザイク化などの不鮮明化処理を施した，一見すると無意味に見える画像をパスワード画像として使用する．正規ユーザにのみ，パスワード画像登録時に不鮮明化処理を行う前のオリジナル画像を見せ，不鮮明化されたパスワード画像をオリジナル画像の持つ意味と結び付けて記憶することができるようにする．これにより，正規ユーザ自身は認証時に表示される不鮮明なパスワード画像を容易に再認することが可能となり，正規ユーザの記憶負荷を小さくできる．一方，オリジナル画像を見ることのできない他のユーザが正規ユーザの認証行為を覗き見たとしても，意味を認識できない不鮮明なパスワード画像を記憶にとどめておくことは困難である．

また，現実には，覗き見などで認証情報が盗まれるほかにも正規ユーザ本人が認証情報を漏洩させる場合が往々にしてある[16),17)]．文字や数字をベースとしたパスワード方式では，パスワードを書き下しておいたものを読まれたり，電子メールや電話などを用いてパスワードを教えたりすることによって，簡単かつ正確に自身のパスワードを他人に伝達することが可能である．組織内や友人間でのソフトウェアのカジュアルコピーなど，正規ユーザ本人からのパスワードやシリアルナンバの漏洩が問題視される場合も多い[14)]．

文献 5) では，ランダムに生成された幾何学模様の人工画像[10)]をパスワード画像として用いることで，正規ユーザ本人から他者にパスワードの特徴を一意に伝達することを難しくしている．しかし，色情報や，円・直線といった形状や位置の情報を言語化して伝えることは比較的容易であり，そこからパスワード画像が漏洩する可能性は大きい．本方式では，モノトーン化やモザイク化を施した，認識が困難な不鮮明化画像をパスワード画像として用いているため，文献 5) に比べて，色や形状の情報を用いてパスワード画像の情報を言語化して他者に伝えることがより困難になっていると考えられる．

本論文では，画像の不鮮明化によって得られるパスワード画像の覗き見や言語による漏洩に対する耐性を調べるため，本方式の基本的なシステムを実装し，各種の基礎実験を行った結果を述べる．

2. 提案方式のコンセプト

画像認証方式では，人間の得意とする画像記憶を用いることで正規ユーザの記憶負荷を軽減させるが，同時に，覗き見をする攻撃者にとっても他者のパスワード画像を記憶する際の負荷が軽減されることになる．また，タッチパネルディスプレイに画像を規則正しくならべて呈示し，パスワード画像にタッチする（PC のディスプレイにおいてはマウスクリック）だけで認証が可能であるような高いユーザビリティは，攻撃者にとっても覗き見をしやすい環境を与えてしまっている．このように，覗き見は画像認証方式の本質に関わ

る攻撃であるといえる．画像をユーザに提示する必要があるため，通常の文字によるパスワードのように入力された文字を「********」と表示して隠すような対策はとれない．また，表示される画像に番号を付けておき，ユーザの手元を隠して，パスワード画像の番号をキー入力することによって認証を行うことも考えられるが，タッチパネル操作のようなユーザビリティは損なわれてしまう．そこで，本論文では，画像認証方式が有する記憶負荷や操作性における優位さを保ちつつ，覗き見攻撃に対する耐性を向上させる方式を検討していく．

本提案方式では，覗き見をする攻撃者にとってパスワード画像の記憶が困難となるように，モザイク化などの不鮮明化処理を施した，一見すると無意味な画像をパスワード画像として使用する．人間は画像の記憶に優れているものの，それは有意味な画像を記憶する場合であり，無意味に見える（意味を言語化できない）画像を記憶することはやはり難しい[13]．したがって，無意味に見える不鮮明なパスワード画像を記憶することは攻撃者にとって困難な作業となる．ただし，完全に無意味な画像を用いると正規ユーザにとっても記憶が困難になってしまう．

そこで本方式では，正規ユーザにのみ，不鮮明なパスワード画像を記憶するための方法を学習させる．すなわち，パスワード画像の登録時に，不鮮明化処理を施す前の有意味なオリジナル画像（写真画像など）を，パスワード画像とともに正規ユーザに提示する．不鮮明化処理を施された画像（パスワード画像）はある程度オリジナル画像の特徴を残しているため，オリジナル画像を見た正規ユーザは，不鮮明なパスワード画像の中にオリジナル画像の持つ意味を見い出すことが可能であり，当該パスワード画像を有意味な画像として認識できるようになる．オリジナル画像を見せるという方法により，正規ユーザにのみ不鮮明なパスワード画像をオリジナル画像と結び付けて記憶させることを実現しており，これは，正規ユーザにのみパスワード画像に対する「スキーマ」を学習させていることに相当する．ここでいう「スキーマ」とは，人間が物事をどのように記憶したかという知識構造を意味する認知心理学用語である．一度スキーマが形成されれば，以後，当該パスワード画像を見た際にオリジナルの意味を再認識することは容易になる[13]．

ここで重要なことは，オリジナル画像を見ていない他のユーザにはパスワード画像からオリジナル画像の意味が類推できない程度に不鮮明化処理を行うようにする（他のユーザには当該パスワード画像に対するスキーマを与えない）ことである．正規ユーザはスキーマを有しているために，認証時に不鮮明なパスワード画像を見れば，当該パスワード画像中にすぐにオリジナル画像の意味を見い出し，再認によってこれを思い出すことができる．よって，正規ユーザは不鮮明なパスワード画像を細部まで記憶しておく必要はなく，記憶の負荷は小さい．一方，攻撃者は他者のパスワード画像に対するスキーマがないために，当該パスワード画像を無意味な画像として認識せざるをえず，たとえパスワード画像（不鮮明化画像）を覗き見たとしても記憶にとどめることが困難になると考えられる．

なお，パスワード画像の覗き見に対する耐性を改善するには，認証画面を毎回ランダムにグループ分けし，パスワード画像が属するグループ番号を返答することで認証を行うようなチャレンジ&レスポンス方式[20]を利用することや，認証画面をいったん全消去した後にパスワード画像が表示されていた位置を指し示したりするなどの，パスワード画像の行方をくらます方法も有効である．しかし，本論文の目的は「攻撃者にパスワード画像を覗き見られても，攻撃者にパスワード画像が漏洩しない」ような画像認証方式を実現することにあるので，パスワード画像の行方をくらます方法については検討の対象外とする．

また，不鮮明化画像に対するオリジナル画像を実際に目にしない限り，正しいスキーマを形成することは困難であることから，正規ユーザが自身のパスワード画像（不鮮明化画像）の特徴を言葉で他者に伝えようとしたとしても，スキーマを有していない他者がそれを理解することは難しいと考えられる．よって本方式は，正規ユーザが自身のパスワード画像の情報を言語化して他者に伝えるというセキュリティホールに対しても，ある程度の効果が期待される．少なくとも，正規ユーザが友人に自身のパスワードを安易に電話で伝えるというような状況は防止することができるだろう．

3. 基本方式：不鮮明化画像を用いる認証方式

本章では，提案方式の基本となる認証方式のプロトタイプシステムを実装し，本基本方式における覗き見攻撃と言葉によるパスワード画像漏洩に対する耐性を評価する基礎実験を行う．

3.1 認証に使用する不鮮明化画像

認証に使用する画像について説明する．基本方式では，多数の写真画像などの有意味なカラー画像 $I(x,y)$（以下，オリジナル画像と記す）と，$I(x,y)$ に対してモザイク化などの不鮮明化処理を施した画像 $O(x,y)$（以下，不鮮明化画像と記す）を使用する．以下に，今

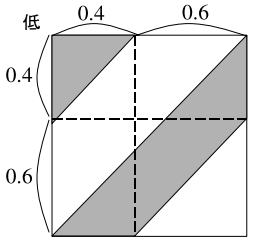

図 1 不鮮明化処理における DCT 係数の変更範囲
Fig. 1 Noise range for the basic scheme.

回のシステムで採用した不鮮明化処理の手順について説明する．

Step0：300×300 ピクセルの 256 色カラー画像 $I(x,y)$ を用意する．

Step1：$I(x,y)$ をモノトーン化した後，ヒストグラム均一化処理をして，明るさおよびコントラストを調整した画像 $I'(x,y)$ を得る．

Step2：$I'(x,y)$ に対し，6×6 ピクセルブロック単位でモザイク化処理を行い，画像 $I''(x,y)$ を得る（各ブロックは，ブロック内の平均輝度で 1 色にぬりつぶされる）．

Step3：$I''(x,y)$ のモザイク処理された各ブロックを 1 画素と見なした画像 $M(k,l)$（50×50 ピクセル）に対して，二次元 DCT 処理を行う．今回は簡単のため画像全体を 1 ブロックとして DCT を行った．

Step4：Step3 で得られた DCT 係数の低周波成分および中〜高周波成分の値にノイズとなるデータを与える．今回のシステムでは，**図 1** におけるグレーの範囲に対応する DCT 係数に，−100〜100 の値をランダムに代入し，DC 成分は 0 とした．その後，IDCT 処理によって画像 $M'(k.l)$ を得る．今回のシステムでは，乱数のシードにつねに同じ値を設定し，同じ画像に対してはつねに同じ不鮮明化画像が作成されるようにした．

Step5：$M'(k,l)$ の 1 画素を 6×6 サイズのブロックに伸長し，元画像の大きさに戻した後，再びヒストグラム均一化の処理を行って画像 $I'''(x,y)$ を得る．

Step6：$I'''(x,y)$ に対して，$I''(x,y)$ を重み $w(0\le w\le1)$ の加重平均によって重ね合わせ処理を行い，画像 $O(x,y)$ を得る．

$$O(x,y)=wI''(x,y)+(1-w)I'''(x,y),\quad \forall(x,y)$$

今回のシステムでは，$w=0.3$ とした．

Step4 における DCT 係数の操作による画像の劣化の程度には画像ごとに大きな差がでるため，Step4 では比較的大きく画像を壊しておき，Step6 の処理によってオリジナル画像の特徴を補完してバランスをとっている．Step4 において各画像に応じて適切な DCT 係数の調整が行えれば，Step6 の処理は必要ない．

上記の手順に従ってオリジナル画像から得られる不鮮明画像の例を**図 3** に示す．図 3 左はオリジナルのカラー画像であり，図 3 右は不鮮明処理後の画像である．不鮮明化画像は，オリジナル画像と比較して，モザイク化や DCT 係数の操作によって大きく情報量が削減されているが，ある程度の特徴が残されていることが見てとれる．

3.2 基本方式の認証手順

今回のシステムにおけるパスワード画像登録および認証の手順は以下のとおりである．

● 登録フェーズ

Step1：認証システムはユーザに複数のオリジナル画像を提示する．

Step2：ユーザはパスワード画像として用いたい画像を選択する．

Step3：認証システムは，ユーザが選択した画像に対応する不鮮明化画像をユーザに提示する．

Step4：ユーザは，オリジナル画像と不鮮明化画像を比較しながら記憶する．

Step5：ユーザが納得すれば，認証システムは当該不鮮明化画像をパスワード画像として登録する．

● 認証フェーズ

Step1：認証システムはユーザに対して，当該ユーザのパスワード画像を含む複数枚の不鮮明化画像をランダムに選び，規則正しく並べて提示する．

Step2：ユーザは，提示された不鮮明化画像の中から，自身のパスワード画像を探し出す．

Step3：ユーザが正しい位置を選択することができれば認証成功とする．

要求される認証強度に応じて，パスワード画像の枚数，認証時に提示される不鮮明化画像の枚数，認証フェーズの繰返し回数（ターン数）などが定められる．不鮮明化画像をパスワード画像として用いることを除くと，根本的には既存の画像認証方式[5)〜9)] と同様の手順である．

3.3 基本方式における覗き見攻撃の実験

基本方式の覗き見攻撃に対する耐性を比較実験によって評価する．比較対象となる認証システムは，基本方式と同様の手順（3.2 節の認証手順）により認証を行うが，不鮮明化画像を使用せず，オリジナル画像そのものをパスワード画像とする認証システムである．これは，既存の画像認証方式を想定したものである．

両システムとも，ユーザのパスワード画像の枚数は 1 枚，認証画面に提示される画像は 2 枚，ターン数は 1 回とした．すなわち，認証は 2 択によって行われる．

また，被験者（攻撃者）には覗き見の直後での成りすましを試みさせた．これらの前提をおいた理由は，攻撃者にとって非常に有利な条件下においてでも，本方式の効果が認められるかどうかを測るためである．

● 実験システム

2 つの認証システムは次の仕様とした．

1) オリジナル画像による認証システム

認証用画像として，4 足哺乳動物のカラー写真画像 90 枚を使用する．これらの画像は，様々な種類の 4 足哺乳動物が背景つきで写されている．認証手順は 3.2 節の手順と同様だが，パスワード画像としてオリジナル画像を使用する．ユーザが記憶すべきパスワード画像は 1 枚のみであり，認証時にユーザに提示される画像は 2 枚である．すなわち，認証に成功するか否かは 2 択であり，ランダムな選択によっても 1/2 の確率で認証に成功する．**図 4** に実際の認証画面の例を示す．なお，以降も含めて，本論文に掲載されるオリジナル画像はインターネット上で収集した画像のうち，著作者が任意の公開および使用を認めているものである．

2) 基本方式の認証システム

3.2 節の認証手順により認証を行う．認証に用いる画像として 90 枚の不鮮明化画像を使用した．この画像は，オリジナル画像による認証システムで使用した 90 枚のカラー写真画像を不鮮明化処理した画像である．ユーザが記憶すべきパスワード画像は 1 枚であり，認証時に表示される画像は 2 枚である．**図 5** に，実際の認証画面の例を示す．

● 実験方法

被験者は，本学男子学生 10 名である．なお，以降のすべての実験において被験者は同じ人物である．各実験システムに対して，実験実施者（正規ユーザ）が認証を通過するところを，被験者（攻撃者）はすぐ隣にいて覗き見る．その直後に，被験者は正規ユーザに成りすまして認証を行う．その際の成功率，および，判断に要した時間（認証画面が表示されてから画像をマウスクリックするまでの時間）を測定する．被験者の覗き見の時間を均一にするために，正規ユーザは認証画面が表示されてから 5 秒後にパスワード画像を選択するようにした．また，1 秒後にパスワード画像を選択する場合についても実験を行った．それぞれの場合において，各被験者につき 5 回ずつ，計 50 回の成りすましを行ってもらった．なお，以降のすべての実験において，システムの方式については事前に十分に説明し，被験者が納得するまで練習を行ってから実験を行った．また，認証時にどちらがパスワード画像か迷った際には少しでも近いと思う方を選択させた．

表 1 オリジナル画像による認証システムの覗き見攻撃の結果
Table 1 Observing attack against original pass-image authentication.

覗き見時間	5 秒	1 秒
成功率	50/50 (100%)	50/50 (100%)
平均回答時間	1.332 秒	1.364 秒

表 2 基本方式の覗き見攻撃の結果
Table 2 Observing attack against unclear pass-image authentication.

覗き見時間	5 秒	1 秒
成功率	46/50 (92%)	45/50 (90%)
平均回答時間	2.655 秒	3.133 秒

● 実験結果

オリジナル画像による認証システムの結果を**表 1** に，基本方式の結果を**表 2** にまとめた．

表 1 の結果から，オリジナル画像を用いた実験では，覗き見時間が 1 秒間であっても，攻撃者はパスワード画像を確実に記憶できていることが分かる．一方，表 2 の基本方式の認証システムでは，表 1 に比べて覗き見攻撃の成功率を下げることができている．また，判断に要する平均回答時間が大きくなっていることからも，不鮮明化処理によって覗き見攻撃がある程度困難化されたことが確認できる．しかし，それでも高い確率で成りすましが可能であり，数秒の覗き見によって，不鮮明化画像であっても攻撃者は非常に高い確率で記憶可能であることが分かった．

実験後，被験者（攻撃者）から聞き取り調査を行ったところ，不鮮明化画像からオリジナル画像が類推できたケースは稀であった．しかし，不鮮明化画像の中の特徴的な一部分を記憶し，それを手掛かりとすることにより認証に成功することができたというコメントが多かった．今回は 2 択のシステムであり，一部分だけの特徴を記憶するだけでも画像間の差異として十分なことが多く，これが原因で成りすまし成功率が大きくなったとみられる．

3.4 基本方式における言葉によるパスワード画像漏洩の実験

基本方式の認証システムにおいて，パスワード画像を言葉で他者に伝えることが可能かどうかの実験を行う．

● 実験システム

3.3 節の実験で用いたものと同じ基本方式のシステムを用いた．

● 実験方法

被験者は，本学男子学生 10 名である．実験実施者

表 3 基本方式のパスワード漏洩実験の結果
Table 3 Leakage of unclear pass-images with words.

成功率	37/50 (74%)
平均回答時間	10.910 秒

（正規ユーザ）が被験者に，パスワード画像に対するオリジナル画像の特徴を言葉によって教える．教える内容は，「動物の種類（例：犬）」「正面か，横向きか」「全身か，一部か」「座っているか，立っているか」といった情報である．たとえば，「猫が横向きに立っている姿の全身が写っている画像」というように教える．その情報を聞いたうえで，被験者が正規ユーザに成りすませるかどうかの実験を行った．各被験者につき 5 回ずつ，計 50 回の成りすましを行ってもらった．なお，どちらがパスワード画像か迷った際には，少しでも近いと思う方を選択させた．

● 実験結果

実験の結果を**表 3** にまとめた．

実験結果は，認証成功率が 74%と，1/2 と比較してやや大きくなった．この理由は，不鮮明化画像においてもオリジナル画像の特徴である輪郭線などがある程度は残されており，明らかな体型の差や，向きの違いを認識できるためと考えられる．特に今回は 2 択であるため，それらしくない方を消去法によって捨てることによっても成りすましが成功する．実験後の聞き取り調査からも，これを裏づけるコメントが多数得られた．

3.5 基本方式における本人認証の実験

スキーマを有している正規ユーザであれば，不鮮明化画像を用いての認証が可能である．しかし，不鮮明化画像はオリジナル画像と比べて情報が大幅に欠落しているため，正規ユーザの本人認証率に少なからず悪影響が及ぶ可能性が否定できない．そこで，基本方式の認証システムにおいて，正規ユーザによる本人認証の実験を行う．

● 実験システム

3.3 節の実験で用いたものと同じ基本方式のシステムを用いた．

● 実験方法

被験者は，本学男子学生 10 名である．被験者に正規ユーザとしてパスワードを登録させ，1 日後と 8 日後に認証を行ってもらい，その際の成功率，および，判断に要した時間（認証画面が表示されてから画像をクリックするまでの時間）を調べた．パスワード登録時には，パスワード画像と対応するオリジナル画像を比較して学習させ，被験者が納得するまで認証の練習を行わせた．1 日後，8 日後の実験では，ともに各被験者につき 5 回ずつ連続で本人認証を行ってもらい，それぞれ計 50 回のデータを得た．パスワード画像は各被験者につき 1 枚であるため，各被験者は 1 日後および 8 日後に，同じパスワードによる認証試行を 5 回ずつ行ったことになる．なお，どちらがパスワード画像か迷った際には，少しでも近いと思う方を選択させた．

表 4 基本方式における本人認証の結果
Table 4 Authentication of the legitimate users with unclear pass-images.

認証実施日	1 日後	8 日後
成功率	50/50 (100%)	50/50 (100%)
平均回答時間	2.239 秒	2.502 秒
1 回目の回答時間の平均	3.201 秒	4.216 秒

● 実験結果

実験の結果を**表 4** にまとめた．

1 日後，8 日後ともに本人認証は完全に成功した．また，5 回続けて行われた認証試行のうち，1 回目の回答時間に比べて，5 回全体の平均回答時間が短くなった．これは，1 回目の認証試行によって，パスワード画像の再認による学習効果があり，2 回目以降の認証試行での反応時間が早まったためであると考えられる．なお，実験後の聞き取り調査からは，各実験の 1 回目の認証試行においてもパスワード画像を再認によって思い出すことができた，あるいは，再認の効果が感じられたというコメントが多く得られた．

3.6 基本方式に関する考察

3.3 節の実験結果より，オリジナル画像を用いる場合よりも，不鮮明化画像を用いる方式では覗き見が困難になっていることが分かる．また，3.4 節の結果から，言葉によるパスワード画像の漏洩抑止にも効果があることも確かめられた（オリジナル画像を用いた方式においては，言葉によってパスワード画像をほぼ 100%伝達できることは自明であろう）．両実験結果とも攻撃の成功率は 1/2 よりも大きかったが，これは攻撃者に非常に有利な条件（2 択のシステム）で実験を行ったことによる結果であると推測される．記憶すべきパスワード画像の枚数や認証フェーズでの選択肢を増やすことにより，攻撃の成功率を十分に下げることができると思われる．

今回の認証に使用した 4 足哺乳動物の画像は，様々な種類の動物を含んでいるため，なかには大きく形態の異なる動物（たとえば，ゾウとウサギ）や，立っている/座っているといった姿勢の違いがあり，不鮮明化処理後の画像においても特徴の違いが現れやすかっ

たということも，攻撃成功率が高くなった原因であると考えられる．認証時の選択肢は 2 択なので，ある一部の特徴を記憶できるだけでも，認証に成功する確率が大きい．たとえば，明確にパスワード画像を選ぶことができなくても，消去法で「それらしくない」と感じる方を捨てることによっても認証が成功する．実際，実験後の聞き取り調査では，オリジナル画像を類推できた例はまれで，一部の特徴のみを利用して成りすましを行ったというコメントが多く得られている．

本人認証に関しては，3.5 節の結果から，不鮮明化画像においても再認の効果が認められた．しかし 3.5 節の実験は，2 択のシステムで記憶すべきパスワード画像が 1 枚という非常に簡単な設定の下における結果である．覗き見耐性に関する考察において「記憶すべきパスワード画像の枚数や認証フェーズでの選択肢を増やすことにより，攻撃の成功率を十分に下げることができると思われる」と述べたが，パスワード画像の枚数や選択肢の増加は正規ユーザの本人認証率および利便性にも影響を与えうる．特に不鮮明化画像を用いる本方式においては，画像が複数になった場合にもパスワード画像を正しく再認可能か否かを調査することは必須といえる．すなわち，本方式の真の有効性を測るためには，認証フェーズにおける選択肢と記憶すべきパスワード画像の枚数を実用レベルにまで増やした場合の本人認証率を測る必要がある．

以上より，本章の基本方式の実験を終えた段階で，1）パスワード画像の枚数と認証時の選択肢を増やした場合に本人認証に悪影響があるかもしれない，2）2 択システムとはいえ，まだ高い確率で覗き見攻撃が成功している，という問題が残った．そこで，以降では，これらの問題に対する追実験を行う．まず 4 章で，問題 1）に対し，記憶すべきパスワード画像の枚数と認証時の選択肢を現実的なレベルに増やした場合の実験を行うことで本人認証への影響を確認する．問題 2）に対しては，5 章において，覗き見攻撃や言葉による漏洩をさらに困難にする方法を提案し，本章と同様の実験を行ってその効果を確かめる．

4. 拡張方式 1：パスワード画像の枚数と認証時の選択肢を増やしての認証

本章では，3 章で使用した基本方式の実験システムを拡張し，記憶すべきパスワード画像と認証フェーズにおける選択肢を現実的なレベルに増やした場合に，本人認証率に悪影響があるかどうかを確かめる．また，同じ設定における覗き見攻撃や言葉による漏洩の成功率についても実験を行う．

本章の実験では，記憶すべきパスワード画像の枚数を 4 枚とし，9 択による認証フェーズを 4 ターン繰り返すシステムを用いる．本システムは「9 枚の画像の中のいずれかがパスワード画像である（または，いずれもパスワード画像でない）」という認証試行を独立に 4 回繰り返すものであり，既存の画像認証方式である文献 5) や文献 6) で行われている検討を参考に，銀行 ATM の数字 4 桁の暗証番号に匹敵した認証（$10^4 = 10000$ の総当り数）を想定している（ただし，今回は「9 枚の画像の中のいずれかがパスワード画像である」という条件で実験を行ったので，正確には 9^4 の総当り数となる）．

4.1 拡張方式 1 における本人認証の実験

● 実験システム

ユーザが登録するパスワード画像は 4 枚であり，認証時に選択肢として表示される画像が 9 枚（横 3 枚 × 縦 3 枚に並べて表示）である以外は 3 章の実験と同じである．認証フェーズでは，9 枚の画像の中からパスワード画像を選択する操作を 4 ターン繰り返し，4 回連続でパスワード画像の選択に成功したときのみ認証成功とする．システムは，1 ターン目の認証画面を用意するにあたり，登録した 4 枚のパスワード画像の中から 1 枚をランダムに選ぶ．同時に，パスワード画像以外の不鮮明化画像の中からランダムに 8 枚を選び，計 9 枚の不鮮明化画像をランダムな配置で表示する．2 ターン目の認証画面においては，残りの 3 枚のパスワード画像の中からランダムに選ばれた 1 枚が，9 枚の不鮮明化画像の中に含まれることになる．3 ターン目は残り 2 枚のパスワード画像のいずれかが，4 ターン目は最後に残ったパスワード画像が選ばれ，それぞれ 9 択の認証画面が構成される．このように，9 枚の不鮮明化画像に中には，いずれかのパスワード画像が必ず 1 枚だけ含まれるようになっている．**図 6** に，実際の認証画面の例を示す．

● 実験方法

登録するパスワード画像枚数，認証フェーズでの選択肢の数および選択の繰返し回数が異なる以外は，3.5 節の本人認証の実験と同様の方法で実験を行った．

● 実験結果

実験の結果を**表 5** に示す．

表 5 中の，「成功率」は 1 日後と 8 日後に各被験者につき 5 回ずつ行った認証試行の全体の成功率を表し，「1 回目の成功率」は，5 回の認証試行のうちの最初の 1 回目のみの認証成功率を表す．また，「ターンごとの平均回答時間」とは，各認証試行において 4 回繰り返すことになる各ターンの認証（9 択の不鮮明化

表 5 拡張方式 1 の本人認証の結果
Table 5 Authentication of the legitimate users with nine-alternative unclear images.

認証実施日	1 日後	8 日後
成功率	50/50 (100%)	49/50 (98%)
1 回目の成功率	10/10 (100%)	10/10 (100%)
ターンごとの平均回答時間	8.194 秒	7.102 秒
ターンごとの平均回答時間（1 回目のみ）	13.124 秒	10.130 秒

画像の中からパスワード画像 1 枚を選択するタスク）1 回に要した時間の平均である．

表 5 より，1 日後，8 日後ともほぼ確実に本人認証に成功していることが分かる．8 日後に 1 回だけ失敗した場合があったが，失敗をした被験者からの聞き取り調査からは，9 択の中にパスワード画像と似た画像が登場したために，うっかり誤選択してしまったが，選択後すぐに失敗に気づいたということであった．また，5 回全体のターンごとの平均回答時間が 1 回目のみのターンごとの平均回答時間よりも短く，3.5 節の 2 択システムの場合と同様，パスワード画像の再認による学習効果も見てとれる．一方，2 択システムの場合と比較すると，選択肢の数が 9 択に増加したことで，パスワード画像を探すための平均時間が 3〜4 倍に増加している．

4.2 拡張方式 1 における覗き見攻撃の実験

4.1 節と同じ実験システムを用いて，パスワード画像の枚数および認証時の選択肢を増やした場合の覗き見攻撃の成功率について実験する．

● 実験システム

4.1 節の実験で用いたものと同じ実験システムを用いる．

● 実験方法

登録するパスワード画像枚数，認証フェーズにおける選択肢の数および選択の繰返し回数が異なる以外は，3.3 節の覗き見攻撃の実験と同様の方法で実験を行った．各被験者は，実験実施者（正規ユーザ）が認証フェーズにおける 9 択の選択を 4 回繰り返して認証を通過するところを覗き見た直後に，正規ユーザへの成りすましを試みる．4.1 節の本人認証の実験から，9 択システムの場合は各ターンの認証行為に平均 10 数秒を要するということが分かったため，本実験における認証時の覗き見時間は 15 秒に設定した．

● 実験結果

実験の結果を**表 6** に示す．

表 6 中の「ターンごとの成功率」は，各認証試行時に 4 回繰り返すことになる各ターンの認証（9 択の不鮮明化画像の中からパスワード画像 1 枚を選択するタスク）を独立にとらえ，その成功率を表したものである．一方，各認証試行において 4 ターンの認証にすべて成功した場合が「成功率」である．

表 6 拡張方式 1 における覗き見攻撃の結果
Table 6 Observing attack against nine-alternative unclear image authentication.

覗き見時間	15 秒
成功率	13/50 (26%)
ターンごとの成功率	125/200 (62.5%)
ターンごとの平均回答時間	18.382 秒

結果より，パスワード画像を 4 枚にし，認証時の選択肢を 9 択にすることによって，覗き見攻撃の成功率を 26%にまで低減できることが確認できる．ただし，ターンごとの成功率が 62.5%であることから単純に計算すると，4 ターン連続で覗き見に成功する確率は $(0.625)^4 \approx 0.153$ と見積もられるはずである．これは，各ターンの認証試行が独立ではなく，1 ターン目で出現しうるパスワード候補は 4 枚であるが，認証に成功するたびに次のターンでのパスワード候補が減っていくので，ターンが進むごとに認証に成功しやすくなっていくためであると考えられる．

4.3 拡張方式 1 における言葉による漏洩の実験

4.1 節と同じ実験システムを用いて，記憶すべきパスワード画像の枚数および認証時の選択肢を増やした場合の言葉によるパスワード画像漏洩の成功率について実験する．

● 実験システム

4.1 節の実験で用いたものと同じ実験システムを用いる．

● 実験方法

登録するパスワード画像枚数，認証フェーズにおける選択肢の数および選択の繰返し回数が異なる以外は，3.4 節の言葉による漏洩の実験と同様の方法で実験を行った．各被験者は，実験実施者（正規ユーザ）から 4 枚のパスワード画像の情報を言葉で教えられた後，正規ユーザへの成りすましを試みる．なお，本実験ではパスワード画像の枚数が 4 枚となって覚えきれなくなるため，被験者が実験実施者から伝えられたパスワード画像の情報を紙にメモすることを許した．

● 実験結果

実験の結果を**表 7** に示す．

パスワード画像を 4 枚にし，認証時の選択肢を 9 択にすることによって，言葉による漏洩の成功率を 0%にすることができた．また，ターンごとの成功率に関しては 30%となった．

表 7 拡張方式 1 の言葉よる漏洩の結果
Table 7 Leakage of pass-images with words in nine-alternative unclear image authentication.

成功率	0/50 (0%)
ターンごとの成功率	60/200 (30%)
ターンごとの平均回答時間	29.049 秒

4.4 拡張方式 1 についての考察

3 章で行った基本方式の実験結果と比較しつつ，拡張方式 1 の実験結果について考察する．

パスワード画像や認証時の選択肢を増やして実験を行ったが，本人認証は 1 日後，8 日後ともほぼ 100%の確率で成功となり，2 択の基本方式と比べても本人認証率は悪化していない．すなわち，パスワード画像や認証時の選択肢をある程度増やしたとしても，正規ユーザにスキーマを与えることによって，比較的容易に複数の不鮮明化画像を記憶および再認させることが可能であることが分かった．ただし，認証にかかる時間については，個々のターンで平均 10 数秒の時間がかかっているため，1 回の認証を完了するまでに正規ユーザにストレスを感じさせるかもしれない．

パスワード画像や認証時の選択肢を増やすという拡張は，本人認証率を悪化させない一方で，覗き見攻撃および言葉による漏洩においては，攻撃成功率を大幅に下げることができている．特に，言葉による漏洩に関しては，成功率を 0%にできるという結果が得られた．

5. 拡張方式 2：重畳不鮮明化画像による認証

3.3 節の基本方式に関する実験結果から，不鮮明化したパスワード画像を用いても，2 択で認証を行うような攻撃者に有利な環境では，高い確率で覗き見攻撃が成功するという結果が得られた．このことから，覗き見をする攻撃者は，図 5 のような不鮮明化画像の中にも何らかの特徴を見い出して自分なりのスキーマを形成し，これを瞬時に記憶することが可能であることが分かる．そこで本章では，そのような攻撃者の画像認識能力を逆手にとる方法によって覗き見やパスワード画像の漏洩の防止効果をさらに高める方法を提案する．

基本方式では，正規ユーザに対して，オリジナル画像を見せることでパスワード画像のスキーマを学習させていた．本拡張方式 2 ではそれに加えて，覗き見をする攻撃者に対して，パスワード画像以外のオリジナル画像を見せて偽のスキーマを与え，パスワード画像の特徴を認識しにくくすることにより，パスワード画像の覗き見をさらに困難にする．

拡張方式 2 では，認証フェーズにおいてユーザに提示する画像として，「ある画像に他の画像を重ね合わせた合成画像」を不鮮明化処理した画像を用いる．このような画像を重畳不鮮明化画像と呼ぶことにする．重畳不鮮明化画像には，2 枚のオリジナル画像の特徴が混在していることになる．一方がパスワード画像のオリジナル画像，他方が囮画像のオリジナル画像である．そして，囮画像の情報がやや認識しやすくなるように 2 枚を重ね合わせたうえで不鮮明化処理を行い，かつ，囮画像のオリジナル画像を覗き見攻撃者に見せるという手順をふむ．この結果，囮画像のオリジナル画像を見た後に重畳不鮮明化画像を見せられた攻撃者には偽のスキーマが形成され，攻撃者はどうしても重畳不鮮明化画像の中の囮画像の特徴を優先して認識してしまい，重畳不鮮明化画像の中のパスワード画像の特徴を認識しにくくなると考えられる．このような結果が起こるのは，人間はある刺激に注目すると，それ以外の刺激を抑制して認識するという性質を持つからである．これは，カクテルパーティ効果[15] と呼ばれる人間の認知の性質として知られている．

なお，正規ユーザには登録時にパスワード画像のオリジナル画像が提示され，正しいスキーマが形成されることは，基本方式と同様である．認証時には，パスワード画像と囮画像から作られた重畳不鮮明化画像および囮画像のオリジナル画像を正規ユーザも目にすることになるが，すでに正しいスキーマが形成されている正規ユーザは，重畳不鮮明化画像の中の囮画像の特徴を意識的に除外して，重畳不鮮明化画像の中のパスワード画像の特徴を認識することが可能である．

本章では拡張方式 2 によるプロトタイプの実験システムを実装し，3 章と同様に覗き見や言葉によるパスワード画像の漏洩に対する耐性を評価する基礎実験を行う．なお，本論文で提案している画像認証方式は不鮮明な画像をパスワード画像として用いるものであるが，拡張方式 2 においては認証試行のたびに異なる囮画像が選ばれ重畳不鮮明化画像が一意に定まらないため，拡張方式 2 を説明するうえでは，登録時に正規ユーザが選んだオリジナル画像をパスワード画像と呼ぶことにする．

5.1 認証に使用する重畳不鮮明化画像

拡張方式 2 では，写真画像などの有意味なカラー画像 2 枚を重ね合わせて作成された画像に対して，モザイク化などの不鮮明化処理を施した画像を用いる．その際，片方の画像を認識しやすくし，もう片方を認識しにくくなるように重み付けをして重ね合わせ処理を行う．前者を「表画像 $I_1(x, y)$」，後者を「裏画像

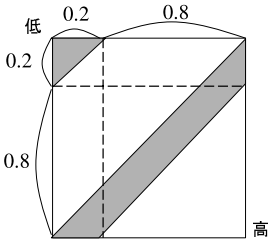

図 2 重畳不鮮明化処理における DCT 係数の変更範囲
Fig. 2 Noise range for the extended scheme.

$I_2(x,y)$」と呼んで区別する．表画像と裏画像を重ね合わせた画像に不鮮明化処理を施すことによって，重畳不鮮明化画像 $O(x,y)$ が得られる．以下に，今回のシステムで用いた，重畳不鮮明化画像の作成手順を示す．

Step0：2 枚の 300×300 ピクセルの 256 色カラー画像 $I_1(x,y)$ と $I_2(x,y)$ を用意する．

Step1：$I_1(x,y)$，$I_2(x,y)$ に対して，モノトーン化処理を行い $I_1'(x,y)$，$I_2'(x,y)$ を得る．

Step2：$I_1'(x,y)$ に対して，$I_2'(x,y)$ を重み $(0 \leq w \leq 1)$ の加重平均による重ね合わせ処理を行い，画像 $D(x,y)$ を得る．

$$D(x,y) = wI_2'(x,y) + (1-w)I_1'(x,y), \quad \forall(x,y)$$

今回のシステムでは，適切な重みで画像の重ね合わせが行われるように，画像の組合せごとに手動で重み w を決定した（w の値はおよそ 0.2〜0.7 の範囲であった）．

Step3：$D(x,y)$ に対してヒストグラム均一化処理をして，明るさおよびコントラストを調整した画像 $D'(x,y)$ を得る．

Step4：$D'(x,y)$ に対し，6×6 ピクセルブロック単位でモザイク化処理を行い，画像 $D''(x,y)$ を作成する．

Step5：$D''(x,y)$ の各モザイク処理ブロックを 1 画素と見なした画像 $M(k,l)$（50×50 ピクセル）に対して，二次元 DCT 処理を行う．今回は簡単のため画像全体を 1 ブロックとして DCT 処理を行った．

Step6：Step5 で得られた DCT 係数の低周波成分および中〜高周波成分の値にノイズとなるデータを加える．今回は，**図 2** のグレーの部分に対応する DCT 係数に，−10〜10 の値をランダムに代入し，DC 成分の値は 0 とした．その後，IDCT 処理をして画像 $M'(k,l)$ を作成する．今回のシステムでは，乱数のシードにつねに同じ値を設定した．

Step7：$M'(k,l)$ の 1 画素を 6×6 サイズのブロックに伸長し，元画像の大きさに戻した後，再びヒストグラム均一化の処理を行って画像 $O(x,y)$ を得る．

上記の方法によって得られる重畳不鮮明化画像の例を**図 7** に示す．図 7 において，左と真中の 2 枚の画像を重ね合わせて不鮮明化処理を行った結果の重畳不鮮明化画像が右の画像である．

5.2 拡張方式 2 の認証手順

拡張方式 2 におけるパスワード画像登録と認証の手順を説明する．

● 登録フェーズ

Step1：認証システムはユーザに複数のオリジナル画像を提示する．

Step2：ユーザはパスワード画像として用いたい画像を選択する．この画像を画像 A とする．

Step3：認証システムは，画像 A に対応する不鮮明化画像をユーザに提示する．

Step4：認証システムは，画像 A とは異なるオリジナル画像をランダムに選んだ（この画像を B とする）うえで，画像 A を裏画像，画像 B を表画像として重畳不鮮明化画像を作成し，これをユーザに表示する．

Step5：ユーザは，画像 A のオリジナル画像と不鮮明化画像，および，画像 A を裏画像として用いた場合の重畳不鮮明化画像の例を比較して記憶することができる（必要ならば Step4 を繰り返して，異なる表画像に対する重畳化不鮮明画像を見ることができる）．

Step6：ユーザが納得すれば，認証システムは画像 A をパスワード画像として登録する．

● 認証フェーズ

Step1：認証システムは，ユーザに対して，当該ユーザのパスワード画像（画像 A）と，画像 A とは異なるオリジナル画像をランダムに選んだ（この画像を C とする）うえで，画像 A を裏画像，画像 C を表画像として重畳不鮮明化画像を作成する．画像 C は認証試行のたびに選び直される．

Step2：認証システムは，画像 A とは異なるオリジナル画像をランダムに 2 枚選んだ（これらの画像を D，E とする）うえで，画像 D を裏画像，画像 E を表画像として重畳不鮮明化画像を作成する．これを繰り返し，複数枚の異なる重畳不鮮明化画像を作成する．なお，すべての重畳不鮮明画像に対する画像 D，画像 E は認証試行のたびに選び直される．

Step3：認証システムは，Step1 で作成した重畳不鮮明化画像と Step2 で作成した複数枚の重畳不鮮明化画像の順序をランダムにシャッフルする．

Step4：認証システムはまず，すべての重畳不鮮明化画像の表画像を，Step3 で決定した順序で規則正しく並べて提示する．

Step5：規定の表示時間が経過した後，認証画面上の表画像はそれぞれに対応する重畳不鮮明化画像に置き換わる．

Step6：ユーザは，提示された重畳不鮮明化画像の中から，自身のパスワード画像が含まれるものを探し出す．

Step7：ユーザが正しい位置を選択することができれば認証成功とする．

要求される認証強度に応じて，パスワード画像の枚数，認証時に提示される重畳不鮮明化画像の枚数，表画像の表示時間，認証フェーズの繰返し回数（ターン数）などが定められる．

本認証手順では，登録フェーズにおいて正規ユーザにパスワード画像（裏画像）に対するスキーマを学習させることによって，正規ユーザがパスワード画像を覚えることを補助することに加え，認証フェーズにおいて囮画像となる表画像を数秒間提示（Step4）した直後に，その表画像の特徴を含む重畳不鮮明化画像を表示（Step5）することにより，攻撃者に偽スキーマを形成させている．なお，認証フェーズにおいては正規ユーザも表画像を目にすることになり，正規ユーザは正しいスキーマと偽スキーマの両方を得ることになる．

5.3 拡張方式 2 における覗き見攻撃の実験

拡張方式 2 の覗き見攻撃に対する耐性を比較実験により評価する．比較対象として，不鮮明化処理を行わない，オリジナル画像そのものを 2 枚重ね合わせた画像を用いて，拡張方式 2 と同様の方法で認証を行うシステムを用いる．

基本方式の評価実験と同様，ユーザのパスワード画像の枚数は 1 枚で，認証画面に提示される画像は 2 枚とした．また，被験者（攻撃者）には覗き見の直後での成りすましを試みさせた．

● 実験システム

2 つの認証システムは次の仕様とした．

1) 重畳オリジナル画像による認証システム

認証用画像として，4 足哺乳動物のカラー写真画像 90 枚を使用する．認証手順は 5.2 節の方式と同様だが，パスワード画像としてオリジナル画像そのものを 2 枚重ね合わせた画像（重畳オリジナル画像）を使用する．その際，重ね合わせの重みは $w = 0.4$ とした．ユーザが記憶すべきパスワード画像は 1 枚，認証時に表示される画像は 2 枚であり，ターン数は 1 回である．認証画面に重畳オリジナル画像が表示される前に，表画像が 3 秒間表示される．被験者にはこの画像を必ず見るようにしてもらった．**図 8** に，実際の認証画面の例を示す．

表 8 重畳オリジナル画像による認証システムの覗き見攻撃の結果
Table 8 Observing attack against overlaid original pass-image authentication.

覗き見時間	15 秒	10 秒	5 秒
成功率	50/50 (100%)	49/50 (98%)	48/50 (96%)
平均回答時間	3.198 秒	3.049 秒	3.355 秒

表 9 拡張方式 2 の覗き見攻撃の結果
Table 9 Observing attack against overlaid unclear pass-image authentication.

覗き見時間	15 秒	10 秒	5 秒
成功率	37/50 (74%)	27/50 (54%)	28/50 (56%)
平均回答時間	5.524 秒	5.997 秒	4.902 秒

2) 拡張方式 2 の認証システム

認証用画像として，4 足哺乳動物のカラー写真画像 10 枚を使用した．この 10 枚の画像の中から表画像と裏画像を 1 枚ずつ選び，重畳不鮮明化画像が生成される．画像数が少ないのは，今回は重畳不鮮明化画像を作成する際の重みを手動で調整したため，多数の画像を用意できなかったためである．重ね合わせのパターンの総数は，同じ画像の重ね合わせを除く 90 通りになる．ユーザが記憶すべきパスワード画像は 1 枚であり，認証時に表示される画像は 2 枚である．重畳オリジナル画像による認証システムと同様，認証画面に重畳不鮮明化画像が表示される前に，表画像が 3 秒間表示される．**図 9** に，実際の認証画面の例を示す．

● 実験方法

3.3 節で行った基本方式の覗き見実験の方法と同様である．ただし本実験においては，認証フェーズで表画像が 3 秒表示された後，重畳不鮮明化画像に切り替わった時点から，パスワード画像を裏画像として含む重畳不鮮明画像を選ぶまでの時間を，認証に要した時間とする．また，重畳不鮮明化画像の中から正解画像を選ぶタスクは正規ユーザにとってもある程度の時間を要求することを鑑み，本実験では，重畳不鮮明化画像の提示から実験実施者（正規ユーザ）が画像を選択する時間（被験者（攻撃者）が覗き見をすることができる時間）を 15 秒間，10 秒間，5 秒間とした．重畳オリジナル画像を用いた認証システムにおける実験も同様に行った．

● 実験結果

実験結果を**表 8**，**表 9** に示す．

重畳オリジナル画像を用いた認証方式では，覗き見の時間にかかわらず，95%以上の攻撃成功率となった．一方，拡張方式 2 では，覗き見時間が 5 秒および 10 秒の場合では，1/2 に近い確率まで成りすましの成功

表 10 拡張方式 2 のパスワード漏洩実験の結果
Table 10 Leakage of overlaid unclear pass-images with words.

成功率	27/50 (54%)
平均回答時間	12.827 秒

率を下げることができ，攻撃者に対してパスワード画像の情報を隠蔽することができた．ただし，覗き見に 15 秒の時間を与えた場合の結果は 74%の成功率となり，パスワード画像の特徴がある程度漏洩している．

実験後の被験者（攻撃者）に対する聞き取り調査からは，被験者の 9 割から，認証時の最初に見せられる表画像によって，その後の重畳不鮮明化画像を見た際の認識に影響があった（偽スキーマにより，まずは重畳不鮮明化画像の中における表画像の特徴に目を奪われる）というコメントが得られた．ただし，覗き見の時間が十分にあれば，認証時の最初に見せられる表画像はパスワード画像ではないという知識を逆手にとって，重畳不鮮明化画像の中において表画像には含まれない特徴を選択して覚える余裕が生まれ，これにより成りすましが可能であったとのことであった．また，重畳不鮮明化画像からパスワード画像が正しく類推できたというコメントは得られなかった．

本実験の結果から，成りすましに弱い 2 択による認証システムにおいてでさえ，正規ユーザが 5 秒〜10 秒程度で認証を完了することができるならば，拡張方式 2 は覗き見攻撃に対して耐性を持つことが期待できる．

5.4 拡張方式 2 における言葉によるパスワード画像漏洩の実験

拡張方式 2 の認証システムにおいて，パスワード画像を言葉で他者に伝えることが可能かどうかの実験を行う．

● 実験システム

5.3 節の実験で用いたものと同じ拡張方式 2 のシステムを用いた．

● 実験方法

被験者は，本学男子学生 10 名である．実験実施者（正規ユーザ）が被験者にパスワード画像の特徴を言葉によって教える．教える情報の内容は 3.4 節の実験と同様であり，被験者がその情報によって正規ユーザに成りすませるかどうかの実験を行った．各被験者につき 5 回ずつ，計 50 回の成りすましを行ってもらった．

● 実験結果

実験の結果を**表 10** にまとめた．

言葉によるパスワード漏洩の成功率は 54%であり，1/2 に近い確率という結果になった．偽スキーマと重畳不鮮明化画像の使用は，言葉によるパスワード漏洩の防止にも大きな効果があることが分かる．

表 11 拡張方式 2 における本人認証の結果
Table 11 Authentication of the legitimate users with overlaid unclear pass-image authentication.

認証実施日	1 日後	8 日後
成功率	47/50 (94%)	44/50 (88%)
平均回答時間	3.345 秒	6.215 秒
1 回目の回答時間の平均	3.878 秒	5.736 秒

5.5 拡張方式 2 における本人認証の実験

拡張方式 2 においては，正規ユーザも認証のたびに表画像と重畳不鮮明化画像を目にするため，偽スキーマの影響が正規ユーザにも及ぶことになる．そこで，拡張方式 2 の認証システムにおける，正規ユーザによる本人認証の実験を行う．

● 実験システム

5.3 節の実験で用いたものと同じ拡張方式 2 のシステムを用いた．

● 実験方法

被験者は，本学男子学生 10 名である．3.5 節の基本方式における実験と同様に，被験者に正規ユーザとしてパスワードを登録させ，1 日後と 8 日後に認証を行ってもらい，認証成功率と判断に要した時間を調べた．パスワード登録時には，パスワード画像と，パスワード画像を裏画像とする重畳不鮮明化画像を比較して学習させ，被験者が納得するまで認証の練習を行わせた．1 日後，8 日後の実験で，ともに各被験者につき 5 回ずつ連続で本人認証を行ってもらい，それぞれ計 50 回のデータを得た．パスワード画像は各被験者につき 1 枚であるが，毎回の認証試行においてはそれぞれ異なる表画像が選ばれ，パスワード画像（裏画像）と重ねられて重畳不鮮明化画像が生成される．どちらがパスワード画像か迷った際には，少しでも近いと思う方を選択させた．

● 実験結果

実験の結果を**表 11** にまとめた．

1 日後，8 日後ともに本人認証は 90%程度の成功率という結果となった．また，5 回続けて行われた認証試行のうち，1 回目の回答時間の平均と 5 回全体の平均回答時間の間に大きな差がなく，1 回目の認証試行におけるパスワード画像の再認による学習効果が見られない．なお，平均回答時間に関しては，8 日後であっても 6 秒程度であった．ただし回答時間の標準偏差を見てみると，1 日後の認証が 2.843 秒，8 日後が 6.233 秒となっており，時間の経過とともにパスワード画像の特徴を思い出すのにかかる時間の個人差が大きくなっている．

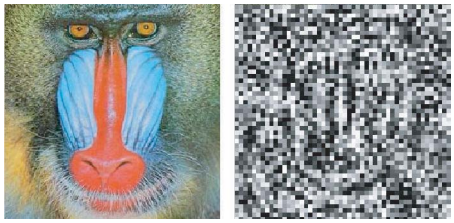

図 3 画像の不鮮明化処理
Fig. 3 An original image and the corresponding unclear image.

図 4 オリジナル画像による認証システムにおける認証画面の例
Fig. 4 Authentication system using original pass-images.

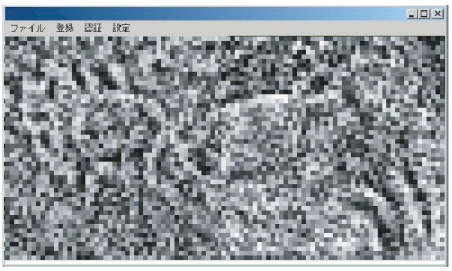

図 5 基本方式の認証システムにおける認証画面の例
Fig. 5 Authentication system using unclear pass-images.

図 6 9 択認証システムにおける認証画面の例
Fig. 6 Authentication system with nine-alternative unclear images.

図 7 重畳不鮮明化画像の例
Fig. 7 Two original images and the corresponding overlaid unclear image.

図 8 重畳オリジナル画像による認証システムにおける認証画面の例
Fig. 8 Authentication system using overlaid original pass-images.

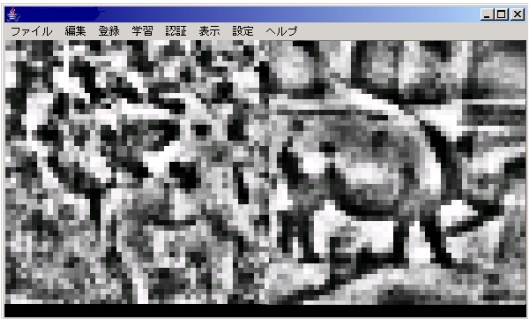

図 9 拡張方式 2 の認証システムにおける認証画面の例
Fig. 9 Authentication system using overlaid unclear pass-images.

5.6 拡張方式 2 に関する考察

拡張方式 2 では，5 秒〜10 秒の覗き見時間であれば，覗き見攻撃をかなり防ぐことができている．被験者には前もって認証方式を十分に説明し，認証時のはじめに表画像として表示されるオリジナル画像はパスワード画像ではないことを教示したため，被験者の多くは重畳不鮮明化画像の中から表画像の特徴ではない部分を記憶すればよいことを知っているのだが，偽スキーマに翻弄されてしまっており，パスワード画像の特徴の漏洩が防止できている．しかし，覗き見時間が長くなると重畳不鮮明化画像の中から表画像の特徴ではない部分を記憶する余裕が生まれ，15 秒の覗き見時間を許した場合にはある程度のパスワード画像の特徴が漏洩する．以上より，本人認証を短時間のうちに

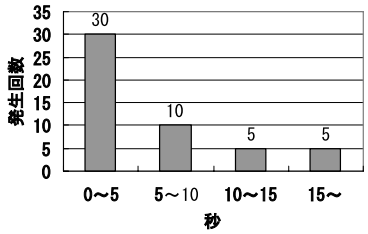

図 10 8 日後の回答時間の分布
Fig. 10 Time required to authentication.

完了させることができるならば，拡張方式 2 の採用により覗き見攻撃に強い認証方式が実現できる可能性がある．また，基本方式と同様に，記憶すべきパスワード画像の枚数や認証フェーズでの選択肢を増やすことにより，攻撃の成功率をさらに下げることができるであろう．

5.5 節の本人認証の実験結果からは，本人認証は 8 日後でも平均 6 秒程度で完了するが，標準偏差が大きく個人差があることが分かった．**図 10** に，8 日後の本人認証における被験者の回答時間を，0〜5 秒，5〜10 秒，10〜15 秒，15 秒以上に分類した度数グラフを示す．図 10 より，全体の 80%において 10 秒以内に認証が完了している．認証に長い時間がかかるケースは正規ユーザにも迷いが生じていることを意味するが，5.5 節の実験結果を解析したところ，確かに，15 秒以内で認証を終えた場合の認証成功率が 90%強であるのに対し，認証に 15 秒以上を要した場合の認証率成功率は 60%程度にまで落ちているという状況であった．よって，認証にある程度の時間を要する場合には認証失敗とするような運用も認証成功率の向上に効果があると思われる．認証にタイムリミットを設けることは，認証画像が覗き見攻撃者の目にさらされる時間を制限することにもつながる．5.3 節の実験結果と合わせて考えるに，認証のタイムリミットは 10 秒程度以下に設定するのがよいと思われる．

拡張方式 2 では覗き見攻撃に対する耐性を強化することができたが，本人認証の成功率が 90%程度に低下してしまうことが大きな問題である．同じパスワード画像（裏画像）を含む重畳不鮮明化画像であっても，重ね合わされるオリジナル画像（表画像）によって，パスワード画像の特徴の出やすさや，特徴が出る部分にばらつきが生じる．表画像と裏画像の動物の形態がよく似ている場合などは，パスワード画像（裏画像）の特徴が表画像の特徴に隠れてしまうこともあった．このように重畳不鮮明化画像においては明確な再認が起こりにくいため，本人認証の成功率が低くなったと考えられる．また，正規ユーザも毎回の認証時に囮となる表画像を見なくてはならないため，表画像を多く見つづけるうちに記憶の衝突が起こり，パスワード画像の記憶が薄れていくということも本人認証率の低下に起因したのではないかと考えられる．本人認証の成功率を改善する方法としては，認証フェーズにおいて，ユーザのパスワード画像に対するヒント（今回の実験の場合，「イヌ」などの動物の種類の名前など）を提示することが有効であろう．5.4 節の実験結果より，言語によってパスワード画像の情報を伝えたとしても攻撃者に手がかりを与えることにはならず，正しいスキーマを持っている正規ユーザにのみヒントが伝わるからである．

表 12 本人拒否率と他人受入率
Table 12 FRR and FAR.

	FRR		FAR	
方式	1 日後	8 日後	覗き見	言葉
基本方式	0%	0%	90%	74%
拡張方式 1	0%	2%	26%	0%
拡張方式 2	6%	12%	54〜74%	54%

6. 検　　討

6.1 本人拒否率と他人受入率

本方式における本人拒否率（FRR）と他人受入率（FAR）について検討する．3 〜 5 章の実験の結果を，FRR と FAR についてまとめたのが **表 12** である．

基本方式において，パスワード画像を 1 枚だけ記憶して，2 択で認証するシステムでは，攻撃者に有利な条件であるということもあり，約 90%という高い確率で覗き見攻撃が成功する．パスワード画像を 4 枚にして，9 択で認証するという拡張方式 1 を採用することにより，FRR を低く保ったまま，FAR を大幅に下げることが可能となる．ただし，まだ覗き見による FAR が 26%ほど残っているため，さらに改善の余地が残る．また，4.4 節で考察したように 1 回の認証に時間がかかるため，正規ユーザの迷いを減らし，パスワード画像を素早く探しあてることができるようにする工夫が必要である．本方式は言葉によるパスワード画像の漏洩を 0%とすることができていることから，言葉によってパスワード画像の特徴を伝えることによって，攻撃者に情報を漏らすことなく，正規ユーザにのみパスワード画像のヒントを与えるというようなことが可能かもしれない．一方，覗き見攻撃への耐性をさらに増すための拡張方式 2 では，2 択システムでありながら，覗き見や言葉による漏洩の FAR を約 1/2 の確率にまで低減できている．ただし，FRR が大きくなる傾向があるため，ユーザビリティという点で問題が残

表 13 本方式と既存方式との認証用画像の比較
Table 13 Comparison with the conventional schemes.

	本方式		既存方式	
	基本/拡張 1	拡張 2	文献 5)	文献 6)
認識・記憶の容易さ	○	△	△	◎
覗き見の困難さ	○	◎	△	×
漏洩の困難さ	○	◎	○	×
推測の困難さ	○	○	○	△

る．拡張方式 2 においては FRR を改善しつつ，画像の枚数やターン数を増加させていくことが今後の課題になる．

6.2 関連方式との比較

本節では，代表的な既存関連研究である文献 5) および文献 6) と本方式について，認証に使用する画像についての比較を行う．文献 5) は，認証システムが自動で生成する幾何学模様の人工画像を使用する．一方，文献 6) では，ユーザ自身のカメラ付き携帯電話などで撮影した写真画像を認証システムに登録して用いる．

これら 3 方式について，認識および記憶の容易さ，覗き見攻撃の困難さ，正規ユーザからの言語によるパスワード画像の漏洩の困難さ，攻撃者によるパスワード画像の推測の困難さについて比較したのが，**表 13** である．

パスワード画像の認識および記憶は，自伝的記憶[13)]と関連づけられた写真画像を用いる文献 6) の方式が最も容易であると考えられる．文献 5) の方式は幾何学模様の人工画像を覚えなければならない．本方式は認識や記憶が困難である不鮮明化画像を用いるが，スキーマを与えることにより，その欠点を補っている．3.5 節の基本方式や 4.1 節の拡張方式 1 の本人認証の実験結果からも，スキーマの効果が裏づけられる．ただし，5.5 節の実験より，拡張方式 2 においては本人認証率が下がることが明らかになっている．

覗き見攻撃と正規ユーザからの言語によるパスワード画像の漏洩に関しては，実際の写真画像をパスワードとして用いる文献 6) の方式の耐性が一番低いと思われる．文献 5) の方式は幾何学模様の人工画像を用いることにより，正規ユーザからの言語によるパスワード画像の漏洩に対する耐性を高めてはいるが，パスワード画像そのものが認証画面に表示される以上，覗き見攻撃に対してはどうしても脆弱になってしまうと考えられる．一方，本方式は不鮮明化画像をパスワード画像に用いることにより，覗き見攻撃および正規ユーザからの言語によるパスワード画像の漏洩に対する耐性を高めることに成功している．特に拡張方式 2 においては，5.3 節，5.4 節の実験結果より，成りすましに弱い 2 択による認証システムにおいてでさえ，十分な効果が得られることが確認できた．

文献 5) の方式，文献 6) の方式，本方式は，「自分の好きな画像を登録しておき，複数の画像の中から正しい画像を選ぶことにより認証する」という認証の根本の方式は同じであるので，intersection 攻撃や educated guess 攻撃などのパスワード画像の推測に対する脆弱度は基本的に同程度であると考えられる☆．ただし，ユーザが自分で撮影した写真そのものをパスワード画像とする文献 6) の方式は，システムが用意した画像の中からパスワード画像を選ぶ方式と比べ，パスワード画像にユーザの趣味や嗜好，行動などの情報が強く現れることになると思われるため，educated guess 攻撃に対する脆弱度が若干大きいと予想される．本方式では，たとえば「正規ユーザはイヌが好きだ」という知識から不正者がパスワード画像（不鮮明化画像）に対する正確なスキーマを得ることは難しい．また，文献 6) の方式では，パスワード画像（自分の撮影した写真）の漏洩がプライバシの漏洩に通じる可能性がある．

6.3 今後の課題

本論文ではパスワード画像として不鮮明化画像を利用した際の効果を調べることに注力し，3 章および 5 章では，最も単純な 2 択のシステムを用いて実験を行った．また，4 章の実験によって，本方式で数字 4 桁の暗証番号に匹敵するレベルの認証システムを実装した場合の覗き見成功率，言葉によるパスワードの漏洩率を評価した．今後はさらに，記憶するパスワード画像の枚数，表示される画像の枚数，表示時間，表示方法などを変えての実験を繰り返し，具体的な認証手順および安全性の指針となるパラメータなどの設定基準を確立していきたい．また，6.2 節で考察したパスワード画像の種類による特徴の差を，実際の実験により確かめたい．

基本方式，拡張方式 1，拡張方式 2 ともに，パスワード画像にオリジナル画像を用いる方式と比べて，認証時にパスワード画像を選ぶまでの時間が増加している．特に拡張方式 2 では，本人認証の成功率そのものが低くなった．本方式における攻撃耐性とユーザビリティの相関を詳細に調べることが必要である．攻撃耐性とユーザビリティには，画像を不鮮明化するアルゴリズムや不鮮明化の度合いも影響してくるだろう．最適な

☆ それぞれの攻撃に対する対処方法に関しては本論文のスコープから外れるため，ここでは言及しない．

不鮮明化画像を自動的に生成する方法についても探索していきたい．

攻撃耐性については，少なくとも以下の検討が必要であると考えている．

- 拡張方式 2 においては，今回の実験では，被験者に認証時に表示される表画像を必ず見てもらうように要求したが，現実には，攻撃者がその間だけ目を閉じておくなどして，偽スキーマの形成を回避する可能性がある．被認証者（が攻撃者であったとしても）に必ず表画像を見せる工夫をするなどの対策が必要であろう．
- 本論文で想定した覗き見攻撃は，近くにいる人間が覗き込むようなカジュアルな方法だが，最近のカメラ付き携帯電話の普及も鑑みると，認証画面をカメラなどで撮影して解析するということもありうる．また，言葉によるパスワード画像の漏洩に関しても，今回の実験で行った方法よりもさらに詳細な情報を伝えたり，紙などにパスワード画像を書いて見せる場合，さらには，パスワード登録時に表示されるオリジナル画像や認証画面をカメラで撮影して見せる場合が考えられる．これらの攻撃への対応を今後検討する必要がある．
- 本方式は，「自分の好きな画像を登録しておき，複数の画像の中から正しい画像を選ぶことにより認証する」という意味では通常の画像認証方式であるので，本質的に，intersection 攻撃や educated guess 攻撃などのパスワード画像の推測に対する脆弱性を有する．これらの攻撃に対する対策を工夫する必要がある．

7. ま と め

本論文では，有意味なオリジナル画像に対してモザイク化をはじめとする不鮮明化処理を施すことで一見して無意味な画像を作成し，それをパスワード画像として用いることで覗き見攻撃への耐性を有する画像認証方式を提案した．正規ユーザのみにパスワード画像に対応するオリジナル画像を見せることで正しいスキーマを学習させることにより，正規ユーザにとっては認識や記憶が容易であるが，覗き見をする攻撃者にとっては「パスワード画像を見ても記憶できない」という状況を作りだしている．また，攻撃者には偽スキーマを与えて翻弄するという方式の提案も行った．本方式のプロトタイプの実験システムを作成して基礎実験を行い，本方式が覗き見攻撃や，正規ユーザからの言葉によるパスワード画像の漏洩に対して耐性を持つことを確かめた．今後，様々な実験を繰り返し，攻撃耐性とユーザビリティを兼ね備える画像認証方式へと昇華させていきたい．

謝辞 株式会社日立製作所システム開発研究所三村昌弘氏と高橋健太氏には，本論文に関する貴重な助言ならびに討議をいただいた．ここに謝意を表す．

参 考 文 献

1) Smith, R.E.（著），稲村 雄（監訳）：認証技術 パスワードから公開鍵まで，オーム社 (2003).
2) Jermyn, I., Mayer, A., Monrose, F., Reiter, M.K. and Rubin, A.D.: The design and analysis of graphical passwords, *Proc. 8th USENIX Security Symposium* (Aug. 1999).
3) Blonder, G.E.: GRAPHICAL PASSWORD, United State Patent 5559961.
4) 鹿島一紀：画像の位置情報による本人認証方式の研究開発画像パスワード GATESCENE（ゲートシーン），情報処理学会コンピュータセキュリティ研究会研究報告, 2000-CSEC-10, pp.121–127 (2000).
5) Dhamija, R. and Perring, A.: Deja Vu: A User Study Using Images for Authentication, *9th USENIX Security Symposium*, pp.45–58 (2002).
6) 高田哲司，小池英樹：あわせ絵：画像登録と利用通知を用いた正候補選択方式による画像認証方式の強化法，情報処理学会論文誌，Vol.44, No.8, pp.2002–2012 (2002).
7) Real User Corporation: PassFace. http://www.realuser.com/cgi-bin/ru.exe/_/homepages/index.htm （2004 年 11 月確認）
8) 有限会社ニーモニックセキュリティ：ニーモニックガード. http://www.mneme.co.jp/neme/neme.html （2004 年 11 月確認）
9) 勝田 亮，平石宏典，溝口文雄：グラフィックパスワードを用いた Web 個人認証システムの設計，情報処理学会コンピュータセキュリティ研究会研究報告 2002-CSEC-16, pp.91–96 (2002).
10) Perring, A. and Song, D.: Hash Visualization: A New Technique to improve Real-World Security, *International Workshop on Cryptographic Techniques and E-Commerce* (*CrypTEC*) (1999).
11) 株式会社 SKR テクノロジー：PPT（Picture Protect Technology）液晶ディスプレイ. http://www.skr-tech.co.jp/2_11.HTML （2004 年 11 月確認）
12) 竹内 啓，西本賢城，佐々木良一：ディスプレイからの視覚的情報漏洩防止システムの開発，情報処理学会コンピュータセキュリティ研究会研究報告 2004-CSEC-25，pp.19–24 (2004).

13) 太田信夫，多鹿秀継（編著）：記憶研究の最前線，北大路書房 (2001).
14) 土屋範久（監修），佐々木良一ほか（編著）：情報セキュリティ事典第7章不正コピー，pp.179–188，共立出版 (2003).
15) 赤木正人：カクテルパーティ効果とそのモデル化，電子情報通信学会誌，Vol.78, No.5, pp.450–453 (1995).
16) ITmedia News 記事：史上最悪のセキュリティホールは，ユーザーのパスワード，2002年5月. http://www.itmedia.co.jp/news/0205/28/ne00_password.html （2004年11月確認）
17) ITmedia Enterprise 記事：従業員はチョコレートバーと引き換えにパスワードを教える—英調査，2004年4月. http://www.itmedia.co.jp/enterprise/0404/21/epn17.html （2004年11月確認）
18) 株式会社東芝：視野角制御フィルタ，東芝レビュー，Vol.59, No.8 (2004). http://www.toshiba.co.jp/tech/review/2004/08/59_08pdf/rd2.pdf （2004年11月確認）
19) Kobara, K. and Imai, H.: Limiting the Visible Space Visual Secret Sharing Schemes and their Application to Human Identification, *Advances in Cryptology* (*ASIACRYPT'96*), LNCS 1163, pp.185–195, Springer-Verlag (1996).
20) 井島裕昭，松本 勉：操作性の良い質問応答型個人認証方式，Proc. SCIS94-13C：1994年暗号と情報セキュリティシンポジウム講演会論文集 (1994).

(平成16年11月29日受付)
(平成17年 6 月 9 日採録)

原田 篤史

平成13年静岡大学情報学部情報科学科卒業．平成15年同大学大学院修士課程修了．現在，同大学院博士課程．情報セキュリティに関する研究に従事．

漁田 武雄

昭和25年生．昭和51年広島大学大学院教育学研究科博士課程後期中退．同年広島大学教育学部助手．昭和55年国立特殊教育総合研究所研究員．昭和57年静岡大学教養部講師．現在，静岡大学情報学部情報社会学科教授．文学博士．人間の記憶の文脈依存機構の解明に関する研究に従事．著書等としては『目撃証言と文脈依存記憶』（現代のエスプリ 350，目撃者の証言：法律と心理学の架け橋，至文堂）等がある．日本心理学会会員，日本認知心理学会理事，日本教育心理学会会員．

水野 忠則（フェロー）

昭和20年生．昭和43年名古屋工業大学経営工学科卒業．同年三菱電機（株）入社．平成5年静岡大学工学部情報知識工学科教授，現在，情報学部情報科学科教授．工学博士．情報ネットワーク，モバイルコンピューティング，放送コンピューティングに関する研究に従事．著書としては『プロトコル言語』（カットシステム），『コンピュータネットワーク概論』（ピアソン・エデュケーション）等がある．電子情報通信学会，IEEE，ACM各会員．当会フェロー．

西垣 正勝（正会員）

平成2年度静岡大学工学部光電機械工学科卒業．平成4年度同大学大学院修士課程修了．平成7年同大学院博士課程修了．日本学術振興会特別研究員（PD）を経て，平成8年静岡大学情報学部助手．平成11年同講師，平成13年同助教授．博士（工学）．情報セキュリティ，ニューラルネットワーク，回路シミュレーション等に関する研究に従事．