情報処理学会論文誌 Vol.61 No.12 1881–1891 (Dec. 2020)

スマートフォンの通話に着目した音声と耳介による個人照合

郷間 愛美[1,a)] 大木 哲史[2] 吉浦 裕[1] 市野 将嗣[1]

受付日 2020年3月7日, 採録日 2020年9月10日

概要：スマートフォンの普及にともない，様々な用途に利用されるようになり高精度な個人照合の必要性が高まっている．それゆえ，身体的特徴や行動的特徴を生体情報として用いるバイオメトリック照合が注目されており，現在ではスマートフォンのログイン時にバイオメトリック照合が利用されている．そこで本研究では，スマートフォンの利用用途の１つである通話に着目し，スマートフォンでの通話時に取得可能な生体情報である音声と，タッチスクリーンから取得可能な耳介を用いたマルチモーダルバイオメトリック照合を提案する．通話中にスマートフォンのタッチスクリーンから耳介データを音声と同時に取得することで，日常での自然な動作を用いて，ユーザに負担をかけることなくシームレスかつ高精度な照合を行うことが可能である．提案手法と音声，耳介の単体での照合の性能を評価，比較し，提案手法の有効性を確認した．また雑音影響下に対しても有効性が確認できた．

キーワード：スマートフォン，音声，耳介，バイオメトリック照合

Personal Identification by Voice and Pinna Focusing on Smartphone Calls

Aimi Goma[1,a)] Tetsushi Ohki[2] Hiroshi Yoshiura[1] Masatsugu Ichino[1]

Received: March 7, 2020, Accepted: September 10, 2020

Abstract: With the spread of smartphones, they are used for various purposes and the need for highly accurate personal verification is increasing. Therefore, biometric matching that uses physical and behavioral characteristics as biometric information is attracting attention, and biometric matching is currently used when logging in to a smartphone. In this study, we focus on calls that are one of the uses of smartphones, and perform multimodal biometric matching using speech, which is biometric information that can be acquired during calls with smartphones, and pinna that can be acquired from a touch screen. We propose acquisition of pinna data simultaneously with voice from the touch screen of the smartphone during a call makes it possible to perform seamless and high-precision collation without burdening the user using natural daily operations. We evaluated and compared the performance of the proposed method with that of the speech and pinna alone, and confirmed the effectiveness of the proposed method. The effectiveness was confirmed under the influence of noise.

Keywords: smartphones, voice, ear, biometric verification

1. はじめに

近年，スマートフォンの普及率が増加している [1]．これにともない，スマートフォンを用いた通販取引やインターネットバンキングなど重要なやり取りを行う機会も増している．そのうえ，スマートフォン内には他人の連絡先などの個人情報，ID，パスワードなどに加えて，クレジットカードや口座番号などの重要な情報も保存されている可能性がある．したがってスマートフォンが悪意のある他人に利用されると，スマートフォン内の情報を悪用される危険性がある．これをふまえてスマートフォンを不正利用から守るために，高精度な個人照合が必要である．

現在スマートフォン上での個人照合には，パスワードや

[1] 電気通信大学
The University of Electro-Communications, Chofu, Tokyo 182–8585, Japan
[2] 静岡大学
Shizuoka University, Hamamatsu, Shizuoka 432–8011, Japan
[a)] goma.a@uec.ac.jp

PIN，パターンなどの知識に基づく照合が主に用いられている．しかしパスワードやPINは，簡単なものを設定している場合も多く，第3者に推測される可能性がある．また電車内などでロックを解除している際に他人に盗み見られ，パスワードなどを知られてしまう恐れも存在する．よって従来のパスワードなどの照合には盗難などの危険が存在する．そこで近年注目されている照合方法の1つにバイオメトリック照合がある．バイオメトリック照合は紛失，盗難などの恐れが低いため，従来の知識に基づく照合と比べて利便性や安全性が高い．

現在，スマートフォン上での主流なバイオメトリック照合は，指紋照合や顔照合である．多くのユーザはスマートフォンに対してロックを施すが，そのスマートフォンに対して1度照合を行うと，その後のユーザが正規のユーザであるかを検証することが少ない．また米国において，スマートフォンユーザの34%がスマートフォンのロックを施さないと記載されている [2]．よって正規のユーザではない人物にスマートフォンを操作され，悪用される危険性がある．この課題の対策として，スマートフォンを利用している最中の行動で照合を行うことが考えられる．

たとえばスマートフォンの主な利用用途の1つである通話の際に得られる音声を利用することで，通話を行うたびに照合を行うことが可能である．通話はスマートフォンの利用用途のなかでも頻度が高く，高頻度に照合を行うことで，スマートフォンのセキュリティを維持することができる．しかし音声のみを用いて照合を行う場合，周囲の雑音などにより照合精度の悪化につながる可能性がある．またスマートフォンは家のなかなどの静かな場所だけでなく，様々な音が飛び交う街中でも頻繁に利用するものであり，スマートフォンでの音声照合を考えるうえで雑音の影響は免れない．加えて，通話時に行う音声照合では録音された音声を利用して，他人になりすまされてしまう場合もある．

そこで本研究では他の生体照合と組み合わせて照合を行う，マルチモーダルバイオメトリック照合を考える．音声だけでなく他の生体情報と組み合わせて照合を行うことで，精度が向上する可能性があり，さらになりすまし対策になりうる．ところが，一般的にマルチモーダルバイオメトリック照合とは，複数の生体情報を取得しなければならず，利便性が損なわれる場合がある．ここでスマートフォンでの通話を考えると，スマートフォンを耳に当てながら会話をするという動作が一般的である．すなわち，通話時の動作でユーザに負担をかけることなく耳介が取得可能である．通話中に耳と接触しているスマートフォンのタッチスクリーンを利用して耳介データを取得することで，日常での自然な動作を用いて，ユーザに負担をかけることなくシームレスかつ高精度な照合を行うことが可能である．またタッチスクリーンから取得する耳介データと音声は同時に取得可能であり，他のモダリティとの組合せと比較して利便性も向上すると考える．

ユーザが企業などからサービスを受ける際に企業などが設置するコールセンターに電話をかけてやり取りをする場合がある．このやり取りにおいて本人確認を求められる場合があり，その際には本人の誕生日などを確認するなど知識に基づく認証が行われる場合がある．このやり取りにおいて通話が用いられている．本論文で提案する通話中の音声とスマートフォンのタッチスクリーンに接触した耳介の情報を利用した照合を通話時の本人確認に利用することで，知識に基づく認証を行わずにユーザに負担をかけることなくシームレスにサービスを受けることが可能となる．

以上より，本論文ではスマートフォンでの耳介と音声を用いた個人照合を提案する．なお本論文は，コンピュータセキュリティシンポジウム2018で発表した内容を発展させたものである [3]．本論文の貢献は，スマートフォンのタッチスクリーンから取得する耳介とマイクから取得する音声を組み合わせる個人照合を提案し，実験，評価を行い，有効性を示したことである．

本論文の構成は以下のとおりである．2章では先行研究，3章では提案，4章では実験，5章では4章の結果，6章では考察，7章ではまとめを記述する．

2. 先行研究

2.1 静電容量画像を用いた個人照合

近年，スマートフォンの静電容量式タッチスクリーンを用いて，手や指などを対象とした照合方法が提案されている．タッチスクリーンの静電容量センサから値を取得し，低解像度のグレースケール画像を作成する．そしてその画像を用いて個人照合を行うという照合方法である．この照合方法ではタッチスクリーンを利用しているため，指紋や虹彩などを用いた照合に比べて特別な照合機器が必要ないという利点が存在する．指紋照合で使われる静電容量センサと比べると，タッチスクリーンに利用されている静電容量センサの解像度は著しく低い．しかし表面積が大きいため，掌などの指紋よりも比較的大きな体の一部を取得可能である点で補うことができる [4]．

Holzら [4] は，LG Nexus5を用いて耳やこぶしなどの5パターンの静電容量画像を取得し，それぞれで照合を行った．取得した 27×15 の低解像度画像の情報強化を行うために，前処理を施したあとにSURF [5] を抽出した．その特徴をテンプレート画像とL2距離で比較し，識別した．耳のみを用いた照合では，被験者12人に対し，本人拒否率が7.8%であった．

Guoら [6] は，Holzらの延長線上の研究を行った．GuoらはNexus5を用いて親指以外の4本の指を用いて照合を行った．取得した画像から550個の特徴を抽出し，そのなかから150個の特徴を選択しSVMを使用した．また，手の水分量の変化による照合精度への影響に関して調査を

行った．加えてHolzらは事後解析のみを行ったため，時間経過による安定性に関しても調査を行った．20人の被験者に対して，FRR4.5%のとき，照合精度は94.5%であった．

Tartzら [7] は，7インチタッチスクリーンに掌と4本の指を押し付けて照合を行った．取得した画像から静電容量値の差を利用し，各指の部分を切り出し，指ごとに特徴量を抽出した．40人の被験者に対して，EERが2.5%であった．

Rilvanら [8] も，耳と指を用いてHolzらと同種の研究を行った．Rilvanらは静電容量センサから取得した値を利用して，耳と指の8 bit，15×27のグレースケール画像を生成した．Rilvanらは新しい特徴量として画像の画素値を利用し，耳の長さと幅，面積の3つの幾何学的特徴を取得した．また15×27のグレースケール画像をを10×10に縮小した画像を作成し，主成分分析を行った．そこで得られた20の主成分と3つの幾何学的特徴に対して，SVMとRF（Random Forest）のそれぞれを利用した．21人の被験者に対して，4本の指とSVMを用いた照合が98.84%と1番の高精度であった．

2.2 耳介と音声のマルチモーダルバイオメトリック照合

岩野ら [9] は，耳介画像と音声のマルチモーダル手法を利用し，話者照合を行った．音声データには，4桁連続数字を静寂な室内で収録し，16 kHz，16 bitで標本化，量子化したものを用いた．学習用音声データには一定の白色雑音を付加し，評価用音声データにはSN比を変えて白色雑音を付加させた．また，耳介画像には右耳正面からデジタルカメラで撮影した解像度720×540の画像を用いた．撮影時の照明条件を一定にするために，フラッシュを使用した．画像に対して位置補正や切り出しを行い，解像度80×80，8 bitのグレースケール画像とし，輪郭協調などの前処理を行った．音声特徴量には，MFCC 12次元，Δ MFCC 12次元，Δ対数パワー1次元の計25次元を用い，数字HMMでモデル化を行った．耳介画像には主成分分析を行い，混合正規分布でモデル化を行った．それぞれのスコアに重み係数をかけた和を融合スコアとし，判定に利用した．性能評価は等誤り率で行い，SN比が30 dBとなるように雑音を付加させた場合が最も高精度であった．

また宮崎ら [10] は，岩野らの研究の発展を行った．岩野らは耳介画像の特徴抽出に主成分分析のみを行ったが，宮崎らは独立成分分析も加えて行った．

3. 音声と耳介を組み合わせたマルチモーダルバイオメトリクスの提案

先行研究 [6], [7], [8] では耳や指などの様々なモダリティに対してタッチスクリーンを用いた照合方法が研究されている．ただしスマートフォンでの照合においてそれらを他のモダリティと組み合わせた照合方法は行われていない．また先行研究 [4] では，タッチスクリーンから取得した複数のモダリティを用いて照合を行っているが，タッチスクリーン以外から取得したモダリティは利用していない．また複数のモダリティを組み合わせて照合を行っているが，単一のモダリティと比較して照合精度が向上していない．さらに先行研究 [9], [10] では，音声と耳介のマルチモーダルバイオメトリック照合を行っているが，スマートフォンでの利用を考えていない．そこでスマートフォンの静電容量式タッチスクリーンから取得した画像を用いた照合と，他のモダリティを組み合わせたマルチモーダルバイオメトリック照合を考える．

本研究では，主なスマートフォンの利用用途である通話に着目し，音声と耳介を用いた照合を提案する．著者らが知る限りでは，音声とタッチスクリーンから取得した耳介を組み合わせたスマートフォンで用いる照合方法がない．スマートフォンでは音声を手軽に入力として得ることができるため，通話の際の照合に利用可能である．またタッチスクリーンの静電容量センサを用いて，低解像度の耳介画像を取得することも可能である．ここで通話の際は耳をスマートフォンに当てながら話すという動作が一般的である．そのため通話の際に利用者に負担をかけることなく，音声と耳介の2つのモダリティを同時に取得することができる．音声と静電容量画像を組み合わせて照合することで，精度と利便性の向上が期待される．なおスマートフォンは様々な場所で利用するものであるため，雑音の影響下においても高精度な照合が期待される．統合手法としては，マルチモーダルバイオメトリック照合の先行研究 [9], [10] で利用されていた重み付け和によるスコア統合のほか，機械学習を用いてスコア統合を行う．

先行研究 [9], [10] では，音声とデジタルカメラで取得した耳介を組み合わせたマルチモーダルを提案している．本論文では，マルチモーダルにする際の手法にスコアレベル統合を利用しているが，先行研究でもよく用いられている手法を用いており，手法自体に新規性はない．本論文の新規性は，音声とタッチスクリーンから取得した耳介を組み合わせて照合するところである．先行研究 [9], [10] ではデジタルカメラを用いて取得した耳介データを用いているが，本提案手法では，タッチスクリーンを用いて取得した耳介データを用いているところが違いとなる．先行研究 [9], [10] ではデジタルカメラを用いて耳介データを取得するので，照合時に耳介をデジタルカメラで撮影する必要がある．提案手法では，通話中の音声データと，通話中に耳と接触しているスマートフォンのタッチスクリーンを利用して耳介データを取得する．そのため，提案手法はデジタルカメラを必要としない．さらに，提案手法により，ユーザにデジタルカメラで耳介を撮影するような負担をかけることなく，日常での自然な動作を用いてシームレスな照合を行うことが可能となるため，提案手法に有効性があ

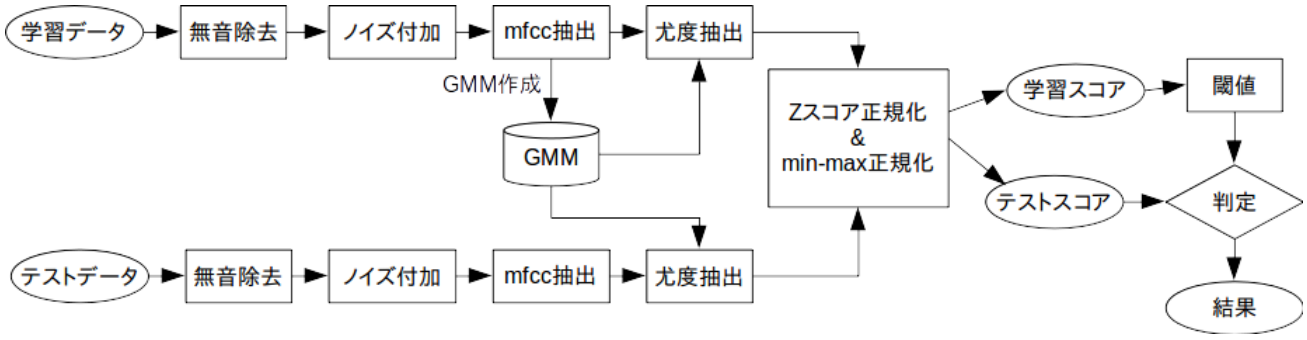


図 1 音声の識別器の構成図
Fig. 1 Diagram of Voice.

ると考える．

4. 実験方法，識別器の概要

本研究では，音声と耳介を組み合わせたマルチモーダルバイオメトリック照合の有効性を示すために，音声，耳介のそれぞれを単体で用いた照合との照合精度の比較を行った．また音声に関しては雑音の種類と SN 比を変化させ，雑音影響下における照合精度の比較を行った．

4.1 音声，耳介データ

被験者 14 人に対して無響室でデータの取得を行った．ATR 文章 [11] の 50 文を 1 文ずつ発話した際の耳介データと音声データを取得した．Nexus5 を用いて右耳の耳介に対応する静電容量値を取得した．耳介データを取得する際には，文献 [12] で公開されていたソースコードを利用した．ただし，このソースコードを使用して耳介データを取得したときに Nexus5 で耳介データの取得以外の操作ができない状態となったため，音声は別の端末で取得する必要があった．そのため，スマートフォンでの照合を考慮し，同時に音声データについては別途 Android スマートフォンのマイクを用いて録音した．

被験者には眼鏡を外してもらい，文章を読み始めるときに耳全体が取得できるように Nexus5 をタッチスクリーンに接触させた．眼鏡を付けていない場合に比べて眼鏡を付けるとタッチスクリーンと耳の接触に変化が生じ，取得される耳介データに変化が生じる可能性がある．今回は音声とタッチスクリーンから取得した耳介を組み合わせた照合の有効性を示すことを目標としており外的な要因で影響を受けていないデータで評価するため，被験者に眼鏡を外してもらった．また，1 文を読み終えるごとに Nexus5 を耳から離し，次の文章を読み始めるときに再度タッチスクリーンに耳を接触するように指示した．これを 50 回繰り返し，1 人につき 50 文の音声と 50 個の耳介データを取得した．取得した音声データは平均して 1 文あたり 6.91 秒であり，耳介データのフレーム数は平均して 1 データあたり 163.7 フレームであった．

本研究では，1 人につき 50 個の音声と耳介のデータのうち，10 個をテストデータ，残りの 40 個を学習データとし，5-fold cross-validation を行った．なお本人同士の照合回数は，1 人あたり 10 個のスコアが 14 人分かつ 5 回分で計 700 回である．本人と他人との照合回数は，1 人あたり 10 個のスコアを 13 人分と照合を行うため，14 人分かつ 5 回分で計 9,100 回である．

4.2 音声の識別器

図 1 に音声の識別器の構成図を示す．

まず始めに，Audacity [13] を利用して音声データの無音部分を削除した．その後，音声データに対して SN 比が 30 dB となるように雑音を付加した．雑音は，電子協騒音データベース [14] の人混み，交差点の 2 種類をそれぞれ用いた．以下，雑音 1 を人混み，雑音 2 を交差点の雑音とする．以降，音声データは，雑音を加えていない音声と，SN 比が 30 dB となるように雑音を加えたの 2 つ音声の計 3 種類の音声を利用した．次に音声特徴量として MFCC 12 次元，Δ MFCC 12 次元，ΔΔ MFCC 12 次元，対数パワー 1 次元，Δ 対数パワー 1 次元，ΔΔ 対数パワー 1 次元計 39 次元の特徴を抽出した．学習データの特徴量を利用し，被験者ごとに混合ガウスモデル（GMM）を作成し，各テストデータの特徴量と GMM との尤度をスコアとした．スコアには，z スコア正規化を行ったあとに min-max 正規化を行った．特徴抽出からスコア算出までは SPTK-3.10 [15] を利用した．

4.3 耳介の識別器

図 2 に耳介の識別器の構成図を示す．

まず前処理として，データの補正を行った．耳介のデータは，静電容量値が 0 より小さい場合は 0 に，255 より大きい場合は 255 に変更し，全静電容量値を 8 bit，0-255 の範囲に補正した．そして，解像度が 24×15 でない場合は，足りない部分を 0 で埋め，全フレームの解像度が 24×15 となるように補正した．次に耳がタッチスクリーンに接触していないフレームを削除した．データの取得では，静電

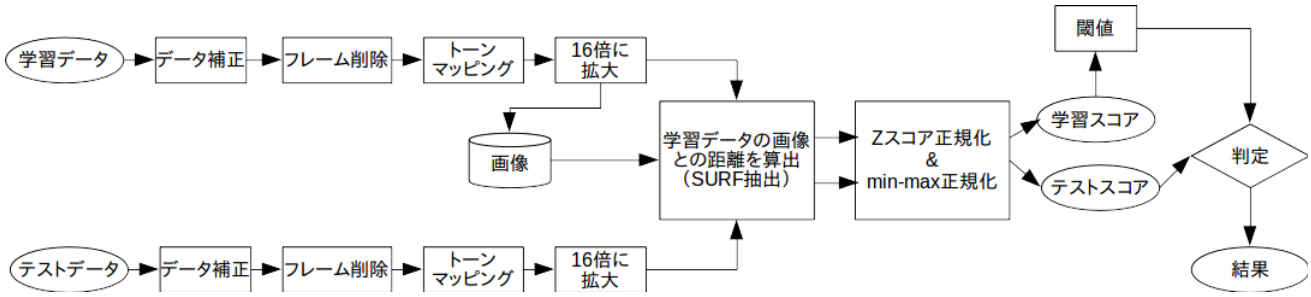


図 2 耳介の識別器の構成図
Fig. 2 Diagram of Ear.

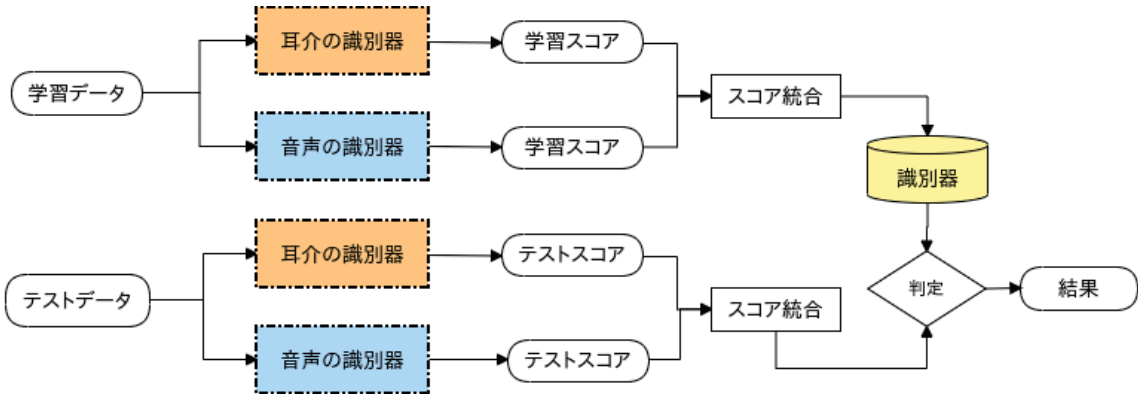


図 3 スコア統合による判定の流れ
Fig. 3 Flow of judgment by integrating scores.

容量値の取得の開始と，実際に耳とタッチスクリーンの接触の開始に時間差が存在した．そのため耳がタッチスクリーンに接触するまでの部分のデータに耳介データは含まれていない．よって耳がタッチスクリーンに接触していないフレームを削除するため，各データ内の 1 フレームごとに平均輝度値を計算し，その値が 10 以下のフレームを削除した．次に低解像度画像を強調するために，データ内の各フレームにトーンマッピング [16], [17] を行った．トーンマッピングとは，局所的なコントラストを保持したまま，画像のダイナミックレンジをディスプレイレンジに圧縮する手法である [18]．最後に 1 データ内の平均フレームを作成し，バイキュービック法を用いて画像の縦サイズを 16 倍，横サイズを 16 倍に拡大した．最終的に 1 文につき 1 枚，1 人につき 50 枚の画像を生成した．

画像特徴量には SURF を用いた．SURF の抽出には OpenCV を用いた．1 人あたり複数個の学習データがある．1 つのテストデータ（1 文から作成される平均フレームをバイキュービック法を用いて拡大した画像）に対して 1 つの学習データ（1 文から作成される平均フレームをバイキュービック法を用いて拡大した画像）との距離を学習データの個数分求めて，それらの距離の平均値をスコアとした．学習のスコア算出も同様であるが，同じ画像どうしの距離算出は除いた．スコアは，z スコア正規化を行ったあとに min-max 正規化を行った．

表 1 実験 A，B，C の音声の条件
Table 1 Experimental conditions of voice.

実験記号	学習データ	テストデータ
A	雑音なし	雑音なし
B	30 dB の雑音 1	30 dB の雑音 1
C	30 dB の雑音 2	30 dB の雑音 2

4.4 音声と耳介の統合

スコア統合による判定の流れを**図 3** に示す．音声と耳介それぞれの識別器から算出したスコアを連結し，2 次元ベクトルで表した．この 2 次元ベクトルの空間において，本人と他人のデータを分離するような識別面を機械学習を用いて設定し，本人か他人か判定した．本研究では，先行研究 [9], [10] で用いられていた重み付け和に加えて，機械学習でよく用いられている SVM（support vector machine）と KNN（k-nearest neighbor）を用いた．

5. 結果

本研究では，音声の条件を**表 1** のように変化させ，3 種類の実験を行った．以下に 3 種類の実験結果を示す．なお，耳介はすべての実験において共通である．また，以下では，音声と耳介を SVM で統合した結果を MM（SVM），KNN で統合した結果を MM（KNN），重み付け和で統合した結果を MM（WS）と表記する．

表 2 実験 A の照合精度の結果
Table 2 Result of Experiment A.

単位：%	Accuracy	Precision	Recall	F1-score	EER
耳介	93.15	52.21	90.43	65.59	11.3
音声	97.23	76.20	89.43	82.21	4.34
MM（SVM）	99.50	98.49	94.57	96.41	1.40
MM（KNN）	99.42	99.28	92.57	95.68	3.71
MM（WS）	99.29	95.85	94.43	94.96	2.23

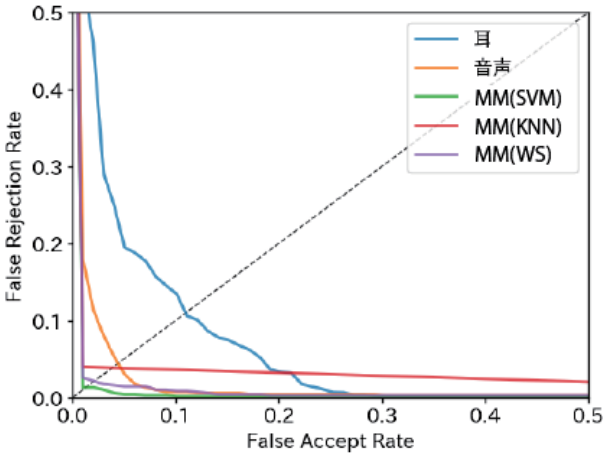


図 4 実験 A の ROC 曲線
Fig. 4 ROC Curve of Experiment A.

5.1 実験 A の実験結果

表 2 に雑音を加えていない音声，耳介をそれぞれ単体で用いた照合とそれらを組み合わせた照合の結果を示す．また**図 4**，**図 5** に ROC 曲線，CMC 曲線を示す．ROC 曲線とは，x 軸，y 軸に FRR，FAR をとり，閾値を変化させたときの結果をプロットした照合性能を表すグラフである [19]．FAR（False Acceptance Rate）とは，他人を誤って受け入れる確率であり，FRR（False Rejection Rate）とは本人を誤って拒否する確率である．また CMC（Cumulative Match Characteristic）曲線とは，順位値を x 軸に，その順位以内での正しい識別率を y 軸に記入した，識別試験の結果を表すグラフである [20]．

表 2 より，Accuracy，Precision，Recall，F1-score，EER のすべてにおいて，単体での照合よりも音声，耳介を組み合わた統合手法の方が精度が向上した．最も高精度であった統合手法は，耳介，音声のそれぞれを用いた照合と比べると Accuracy ではおよそ 6.35%，2.27%，Precision ではおよそ 47.07%，23.08%向上した．また EER では耳介，音声のそれぞれを用いた照合と比較するとおよそ 9.90%，2.94%向上した．さらに図 4 からは単体での照合よりも統合手法のほうが精度が向上していることが確認でき，図 5 では，音声，MM（SVM），MM（WS）の識別率が高いことが確認できた．

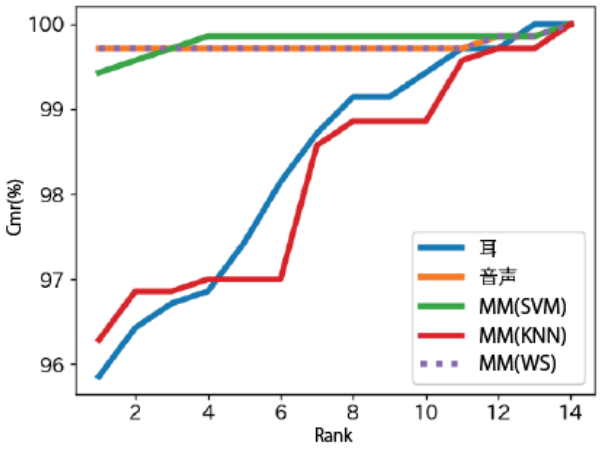


図 5 実験 A の CMC 曲線
Fig. 5 CMC Curve of Experiment A.

表 3 実験 B の照合精度の結果
Table 3 Result of Experiment B.

単位：%	Accuracy	Precision	Recall	F1-score	EER
耳介	93.15	52.21	90.43	65.59	11.30
音声（雑音 1）	97.81	81.57	89.86	85.43	3.41
MM（SVM）	99.53	98.66	94.86	96.63	1.33
MM（KNN）	99.42	98.98	92.86	95.70	3.36
MM（WS）	99.29	95.85	94.43	94.96	1.61

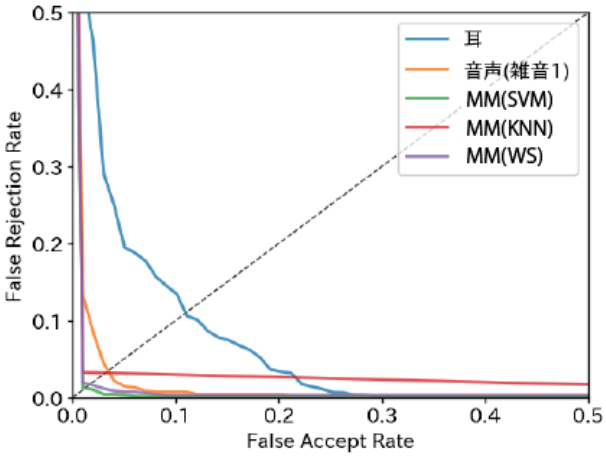


図 6 実験 B の ROC 曲線
Fig. 6 ROC Curve of Experiment B.

5.2 実験 B の結果

表 3 に，雑音 1 を付加した音声と耳介を組み合わせた照合の結果を示す．また**図 6**，**図 7** に ROC 曲線，CMC 曲線を示す．

表 3 より，Accuracy，Precision，Recall，F1-score，EER のすべてにおいて，単体での照合よりも音声，耳介を組み合わた統合手法の方が精度が向上した．最も高精度であった統合手法は，耳介，音声のそれぞれを用いた照合と比べると Accuracy ではおよそ 6.38%，1.72%，Precision ではおよそ 46.77%，17.41%向上した．また EER では耳介，

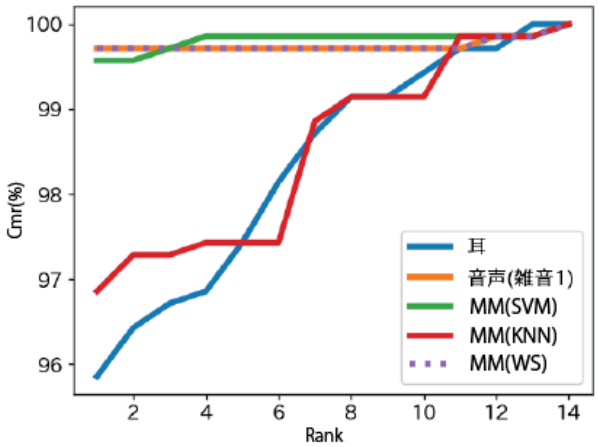


図 7 実験 B の CMC 曲線
Fig. 7 CMC Curve of Experiment B.

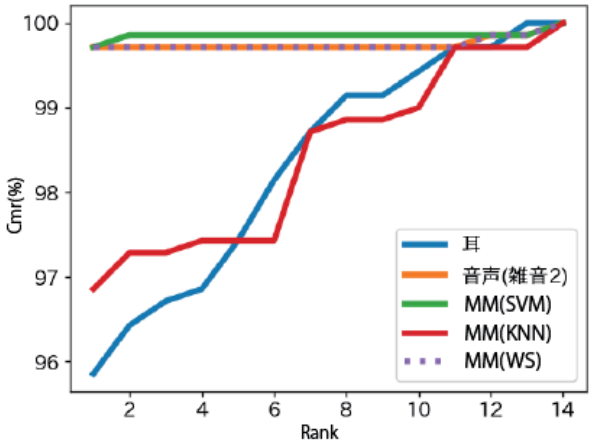


図 9 実験 C の CMC 曲線
Fig. 9 CMC Curve of Experiment C.

表 4 実験 C の照合精度の結果
Table 4 Result of Experiment C.

単位：%	Accuracy	Precision	Recall	F1-score	EER
耳介	93.15	52.21	90.43	65.59	11.30
音声（雑音 2）	98.15	84.97	90.14	87.43	3.05
MM（SVM）	99.58	99.71	94.43	96.94	1.01
MM（KNN）	99.54	99.57	94.00	96.58	3.36
MM（WS）	99.45	97.27	95.14	96.09	1.51

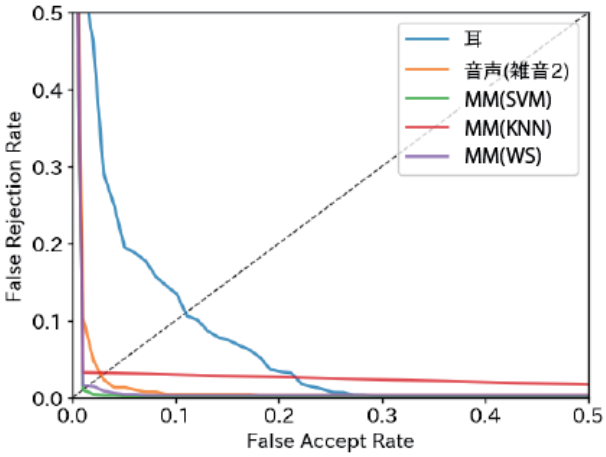


図 8 実験 C の ROC 曲線
Fig. 8 ROC Curve of Experiment C.

音声のそれぞれを用いた照合と比較するとおよそ 9.97%，2.08%向上した．さらに図 6 からは，単体での照合よりも統合手法のほうが精度が向上していることが確認でき，図 7 では，音声，MM（SVM），MM（WS）の識別率が高いことが確認できた．

5.3 実験 C の実験結果

表 4 に，雑音 2 を付加した音声と耳介を組み合わせた照合の結果を示す．また**図 8**，**図 9** に ROC 曲線，CMC 曲線を示す．

表 4 より，Accuracy，Precision，Recall，F1-score，EER のすべてにおいて，単体での照合よりも音声，耳介を組み合わた統合手法の方が精度が向上した．最も高精度であった統合手法は，耳介，音声のそれぞれを用いた照合と比べると Accuracy ではおよそ 6.43%，1.43%，Precision では 47.50%，14.74%向上した．また EER は耳介，音声のそれぞれを用いた照合と比較するとおよそ 10.29%，2.04%向上した．さらに図 4 からも単体での照合よりも統合手法のほうが精度が向上していることが確認でき，図 5 では，音声，MM（SVM），MM（WS）の識別率が高いことが確認できた．

6. 考察

音声のみ，耳介のみを用いた照合と比較して，提案手法の有効性を考察する．加えて，提案手法はスマートフォンで利用するマルチモーダルバイオメトリック照合である．よってスマートフォンでの実用を考慮して，学習データ数，雑音の影響についてそれぞれ考察する．

6.1 提案手法の有効性

図 10 に 5.1 節の音声スコアのヒストグラム，**図 11** に 5.1 節の耳介スコアのヒストグラム，**図 12** に 5.1 節の散布図を示す．図 10，図 11 を見ると本人と他人の分布が重なってることが分かる．よって照合の際に誤りが生じる．

しかし図 12 をみると本人と他人の分布の重なりが小さくなっており，耳介のみ，音声のみと比較して耳介と音声を組み合わせることで，照合の誤りが減少している．したがってスコア分布からも提案手法は耳介のみ，音声のみを用いた照合と比較して高精度な照合であることが分かる．

6.2 学習データ数の変化による精度への影響

4 章では 1 人あたり 50 個のデータのうち 40 個学習データとして用いたが，スマートフォンでの利用を考えると，

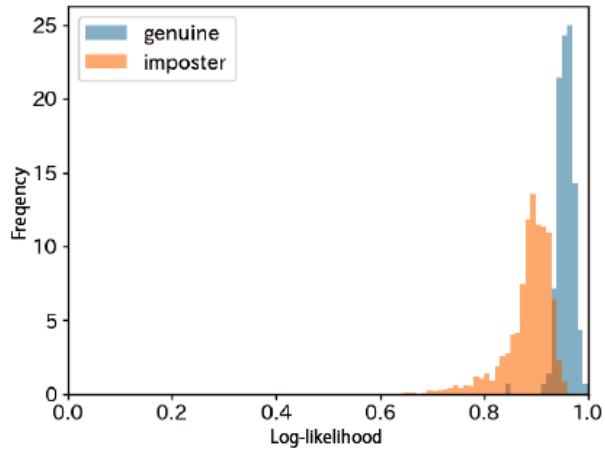


図 10 音声のヒストグラム
Fig. 10 Histogtram of voice.

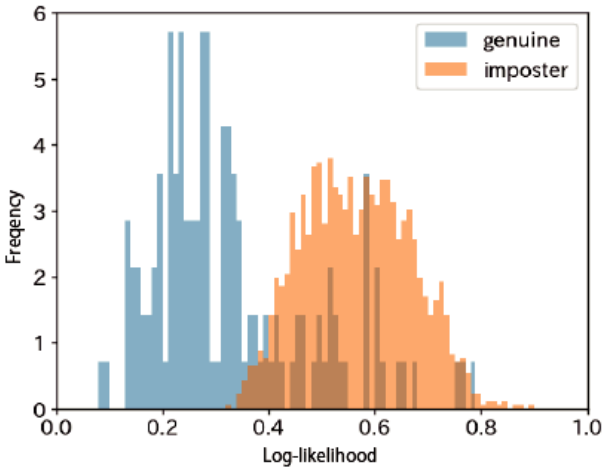


図 11 耳介のヒストグラム
Fig. 11 Histogtram of ear.

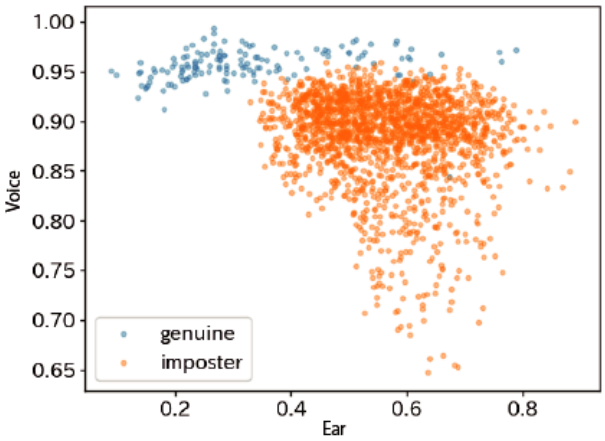


図 12 音声のスコアと耳介のスコアの散布図
Fig. 12 Scatter plot of voice score and ear score.

生体情報の登録回数が多いほどユーザへの負担が大きい．よって学習データ数が少ないほどユーザへの負担が少なく，より実用的な照合方法であると考えられる．そのため学習データ数を 3，5，10，20，30，40 個と変化させ，各照合精度の変化を評価した．なお，テストデータは 4 章と同様に 10 個であり，5-fold cross-validation を行った．各学習データは，テストデータを除いたデータのなかからランダムに選定した．音声，耳介のそれぞれを単体で用いた照合と，音声と耳介を組み合わせた照合の Accuracy の変化を図 **13** に示す．

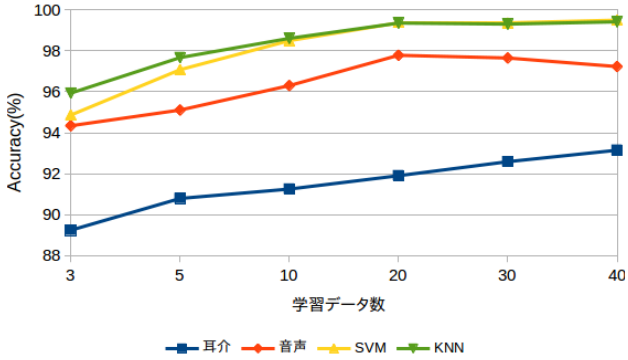


図 13 学習データ数と照合精度の変化
Fig. 13 The Change of the Number of Training data and Accuracy.

表 5 実験 a1，a2，b，c の音声の条件
Table 5 Experimental conditions of voice.

実験記号	学習データ	テストデータ
a1	雑音なし	5，10，20，30 dB の雑音 1
a2	雑音なし	5，10，20，30 dB の雑音 2
b	30 dB の雑音 1	5，10，20，30 dB の雑音 1
c	30 dB の雑音 2	5，10，20，30 dB の雑音 2

図 13 より，学習データが多いほど高精度であるという傾向が確認できる．また，学習データ数にかかわらず，音声と耳介を組み合わせて照合することで，音声，耳介の単体での照合よりも精度が向上している．よって提案手法は学習データ数の変化にかかわらず，単体での照合よりも高精度な照合を行うことが可能である．さらに 1 人あたり 40 個の学習データ数を用いた場合の音声の精度と，1 人あたり 5 個の学習データ数を用いた場合の提案手法の精度を比較すると，およそ同程度である．よって提案手法によって学習データの数を削減することが可能であり，ユーザの負担が少なくなる．

6.3 雑音の影響

4 章では，雑音が比較的小さい環境を想定し，雑音のない音声や SN 比が 30 dB となるように雑音を付加した音声を用いた．しかしスマートフォンは様々な環境下で利用するため，本実験のような雑音が小さい環境だけでなく，雑音が大きい環境で照合を行う場合もある．そこで，雑音の大きさの変化による照合精度の比較を行った．**表 5** に，実験ごとの音声の雑音に関する条件をまとめる．音声は表 5 にある SN 比になるように雑音を付加し，耳介はすべて共

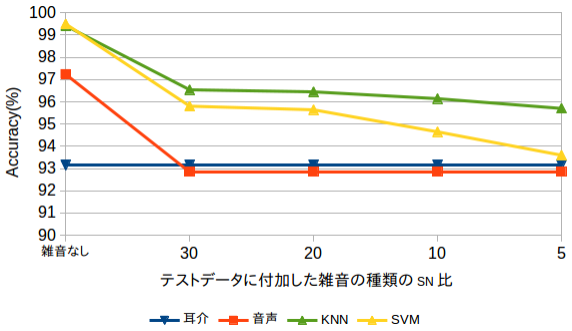


図 14 雑音の大きさによる照合精度への影響（実験 a1）

Fig. 14 Effect of noise size on authentication accuracy (Experiment a1).

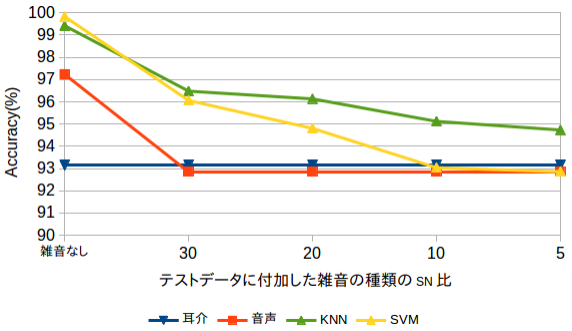


図 15 雑音の大きさによる照合精度への影響（実験 a2）

Fig. 15 Effect of noise size on authentication accuracy (Experiment a2).

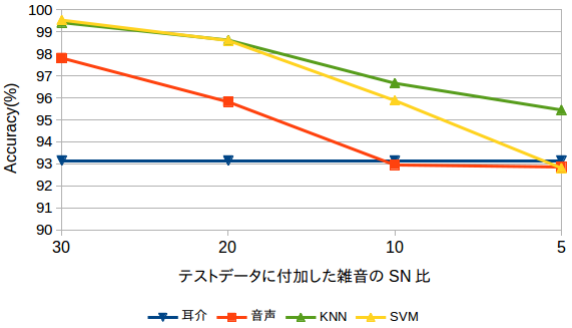


図 16 雑音の大きさによる照合精度への影響（実験 b）

Fig. 16 Effect of noise size on authentication accuracy (Experiment b).

通である．雑音の大きさについては，音声データを利用している先行研究 [9], [21] などで SN 比が 5 dB から 30 dB となるように雑音を付加した音声データが用いられているので，本研究においても SN 比が 5 dB から 30 dB となるように雑音を付加した音声データを利用した．

図 14，**図 15**，**図 16**，**図 17** に実験 a1，a2，b，c の結果を示す．

図 14，図 15，図 16，図 17 より，単体での照合よりも統合手法の方が高精度であり，雑音影響下において精度の改善が見られた．なお雑音影響下において最も高精度な統合手法は，MM（KNN）であった．また本実験では，スマートフォンは様々な場所で利用することを想定して 2 種類の雑音を用いたが，どちらの雑音を用いても同傾向の結果が得られた．よって提案したマルチモーダルバイオメトリック照合は，様々な場所での照合において，単体での照合よりも雑音の影響を低減することが可能である．

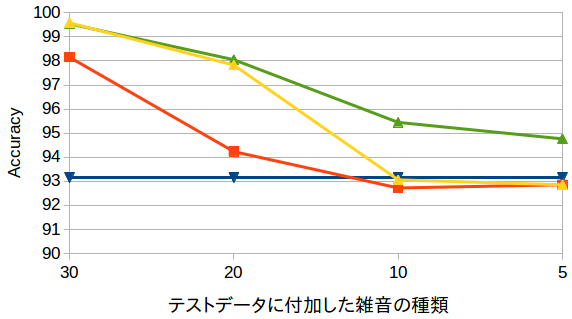


図 17 雑音の大きさによる照合精度への影響（実験 c）

Fig. 17 Effect of noise size on authentication accuracy (Experiment c).

今回の実験では，表 2 と，表 3 および表 4 を見比べると，雑音がある場合（表 3，表 4）のほうが雑音がない場合（表 2）よりも，音声のみの精度評価が高くなっている．表 2 は，雑音を付加しないデータでしきい値設定をした結果である．表 3，表 4 は，雑音を付加したデータでしきい値設定をした結果である．しきい値の値を確認すると雑音を付加したデータで設定したしきい値は雑音を付加しないデータで設定したしきい値よりも大きくなり，スコア分布を確認するとより他人データを他人と判定しやすくなっていた．そのため，雑音がある場合（表 3，表 4）のほうが雑音がない場合（表 2）よりも，音声のみの精度評価が高くなっていたと考えられる．

図 14〜図 17 において，SN 比によらず音声のみが 93%近辺にあり，なおかつ変動しない結果となった．図 14，図 15 の実験では雑音を付加しないデータで GMM を作成した．雑音を加えるとスコアが小さくなり，30 dB の時点で大部分の他人データのスコアがしきい値よりも下回るため，大部分の他人データを他人と判定している．さらに雑音を大きくしてもスコアがしきい値を上回る他人データが増加しないため精度が変動しない結果になったと考えられる．図 16，図 17 の実験では 30 dB の音声データで GMM を作成した．雑音を加えるにつれて他人データのスコアがしきい値よりも下回る傾向にあった．10 dB の時点で大部分の他人データを他人と判定しており，さらに雑音を大きくしてもスコアがしきい値を上回る他人データが増加しないため精度が変動しない結果になったと考えられる．

7. まとめ

本論文では，スマートフォン上で利用する，ユーザへの負担が小さくかつ高精度な個人照合として，音声とスマートフォンのタッチスクリーンから取得可能な耳介を用いた

マルチモーダルバイオメトリック照合を提案した．またその性能を評価し，音声，耳介を単体で用いた照合と比較し，提案手法の有効性を確認した．加えて雑音影響下においても有効であることを確認した．

今後の課題としては，まず耳介，音声のスコア算出の改善，検討や，本研究で用いたSVM，KNN，重み付け和以外の統合手法の検討があげられる．さらに本研究では音声の雑音に対する評価は行ったが，耳介のノイズに関する評価は行っていない．したがってタッチスクリーンから取得した耳介データに関するノイズの調査，および照合精度への影響の評価もあげられる．実際の用途を考えた際に，耳介の一部のみがディスプレイに接触すると考えられる．耳介の情報が欠けることによる影響を調べ，対応策を検討する必要がある．今回の実験では，被験者に眼鏡を外してもらい耳介情報を取得した．眼鏡を付けていない場合に比べて眼鏡を付けるとタッチスクリーンと耳の接触に変化が生じ，取得される耳介データに変化が生じる可能性がある．この影響を調べる必要がある．

参考文献

[1] 総務省：平成29年版情報通信白書｜情報通信機器の普及状況，総務省（オンライン），入手先 ⟨http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/html/nc262110.html⟩（参照 2019-10-22）．

[2] Mahbub, U., Sarkar, S., Patel, V., et al.: Active user authentication for smartphones: A challenge data set and benchmark Results, *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems, BTAS 2016*, IEEE (2016).

[3] 郷間愛美，大木哲史，吉浦裕ほか：スマートフォンにおける音声とタッチスクリーンから取得した耳介を用いた個人認証，コンピュータセキュリティシンポジウム 2018 (2018).

[4] Holz, C., Buthpitiya, S. and Knaust, M.: Bodyprint: Biometric User Identification on Mobile Devices Using the Capacitive Touchscreen to Scan Body Parts, *CHI '15 Proc. 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp.3011–3014, ACM (2015).

[5] Bay, H., ESS, A., Tuytelaars, T., et al.: Speeded-Up Robust Features (SURF), *Computer Vision and Image Understanding*, Vol.110, No.3, pp.346–359 (2008).

[6] Guo, A., Xiao, R. and Harrison, C.: CapAuth: Identifying and Differentiating User Handprints on Commodity Capacitive Touchscreens, *ITS '15 Proc. 2015 International Conference on Interactive Tabletops & Surfaces*, pp.59–62, ACM (2015).

[7] Tartz, R. and Gooding, T.: Hand Biometrics Using Capacitive Touchscreens, *UIST '15 Adjunct Adjunct Proc. 28th Annual ACM Symposium on User Interface Software & Technology*, pp.67–68, ACM (2015).

[8] Rilvan, M., Lacy, K., Hossain, M., et al.: User authentication and identification on smartphones by incorporating capacitive touchscreen, *2016 IEEE 35th International Performance Computing and Communications Conference* (*IPCCC*), IEEE (2016).

[9] 岩野公司，広瀬智治，上林英悟ほか：音声と耳介画像を用いたマルチモーダル話者照合，日本音響学会2003年春季公講演論文集，No.3-3-3, pp.109–110 (2003).

[10] 宮崎太郎，浅見太一，岩野公司ほか：音声と耳介画像を用いたマルチモーダル話者照合の高精度化，日本音響学会2004年秋季公講演論文集，No.2-4-7, pp.99–100 (2004).

[11] ATR503文の全文 - satoru.netの自由帳，入手先 ⟨https://satoru-net.hateblo.jp/entry/20151030/1446184756⟩（参照 2019-10-22）．

[12] Zinda, I.: RainCheck, available from ⟨http://isaaczinda.com/raincheck/index.html⟩ (accessed 2019-10-22).

[13] Audacity ®: Free, open source, cross-platform audio software for multi-track recording and editing, available from ⟨https://www.audacityteam.org/⟩ (accessed 2019-10-22).

[14] 電子協騒音データベース：SHACHI_ Language Resource Metadata Database，入手先 ⟨http://shachi.org/resources/4313?ln=jpn⟩（参照 2019-10-22）．

[15] Speech Signal Processing Toolkit (SPTK), available from ⟨https://sourceforge.net/projects/sp-tk/files/SPTK/SPTK-3.10/⟩ (accessed 2019-10-22).

[16] t-pot『Tone mapping』, available from ⟨http://t-pot.com/program/123_ToneMapping/index.html⟩(accessed 2019-10-22).

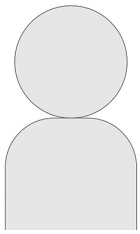
[17] 山内拓也，三上俊彰，宮田公佳ほか：高ダイナミックレンジ画像のための注視領域情報を用いたトーンマッピング手法の評価，日本写真学会誌，Vol.75, No.1, pp.87–96 (2012).

[18] 奥田正浩：HDR画像～色空間から符号化まで～，入手先 ⟨https://www.jstage.jst.go.jp/article/itej/64/3/64_3_299/_pdf⟩（参照 2019-10-22）．

[19] 小松尚久：バイオメトリクスのおはなし，日本規格協会 (2008).

[20] JIS X 8101-1_2010 情報技術—バイオメトリック性能試験及び報告—第1部：原則及び枠組み，入手先 ⟨http://kikakurui.com/x8/X8101-1-2010-01.html⟩（参照 2019-10-22）．

[21] 望月紫穂野，塩田さやか，貴家仁志：話者照合のための音素情報を考慮したポップノイズ検出法による声の生体検知，電子情報通信学会論文誌 D，Vol.J101-D, No.3, pp.588–596 (2018).



郷間 愛美

2020年電気通信大学情報理工学研究科博士前期課程修了．

大木 哲史 （正会員）

2002 年早稲田大学理工学部電子情報通信学科卒業．2004 年同大学大学院理工学研究科電子・情報通信学専攻修士課程修了．2010 年早稲田大学理工学術院情報・ネットワーク専攻博士（工学）取得．2010 年早稲田大学理工学総合研究所次席研究員，2013 年産業技術総合研究所特別研究員を経て 2017 年静岡大学大学院総合科学技術研究科講師，2020 年同大学准教授．情報セキュリティ全般，特に個人認証を中心としたネットワークセキュリティに関する研究に従事．電子情報通信学会会員．

吉浦 裕 （正会員）

1981 年東京大学理学部情報科学科卒業．日立製作所を経て，2003 年電気通信大学勤務．現在，情報理工学研究科教授．情報セキュリティ，プライバシー保護の研究に従事．博士（理学）．日立製作所社長技術賞（2000 年），情報処理学会論文賞（2005 年，2011 年），システム制御情報学会産業技術賞（2005 年），IEEE IIH-MSP best paper award（2006），日本セキュリティ・マネジメント学会論文賞（2010 年，2016 年，2017 年），IFIP I3E best paper award（2016）等受賞．電子情報通信学会，日本セキュリティ・マネジメント学会，人工知能学会，システム制御情報学会，IEEE 各会員．本会フェロー．

市野 将嗣 （正会員）

2003 年早稲田大学理工学部電子・情報通信学科卒業．2008 年同大学大学院理工学研究科博士課程修了．2007 年日本学術振興会特別研究員．2009 年早稲田大学大学院基幹理工学研究科研究助手．2010 年同大学メディアネットワークセンター助手．2011 年電気通信大学大学院情報理工学研究科助教．2016 年同大学大学院情報理工学研究科准教授．バイオメトリクス，ネットワークセキュリティに関する研究に従事．博士（工学）．電子情報通信学会会員．