

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

郷間 愛美^{1,a)} 大木 哲史² 吉浦 裕¹ 市野 将嗣¹

概要: スマートフォンの普及に伴い、様々な用途に利用されるようになり高精度な個人認証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OMETリック認証が注目されており、現在ではスマートフォンのログイン時にバイOMETリック認証が利用されている。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OMETリック認証を提案する。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、ユーザーに負担をかけることなくシームレスかつ高精度な認証を行うことが可能である。提案手法と音声、耳介の単体での認証の性能を評価、比較し、提案手法の有効性を確認した。また雑音影響下や、取得時期に差のあるデータ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, マルチモーダルバイOMETリック認証

Personal Authentication on Amartphone Using Voice and Ear Obtained from the Capacitive Touchscreen

AIMI GOMA^{1,a)} TETSUSHI OHKI² HIROSHI YOSHIURA¹ MASATSUGU ICHINO¹

Abstract: With the spread of smartphones, they are used for various purposes, and the need for highly accurate personal authentication is increasing. Therefore, biometric authentication attracts attention, and now biometric authentication is used at smartphone login. In this research, we focus on calls, which is one of applications for smartphones, and investigate the effectiveness of multimodal biometric authentication. By acquiring the ear from the capacitive touchscreen of the smartphone simultaneously with the voice during the call, it is possible to perform seamless and highly accurate authentication without imposing a burden on the user by using natural actions in daily life. We evaluated and compared the performance of the proposed method, voice and ear authenticaton respectively, and confirmed the effectiveness of the proposed method. It was also confirmed that it is effective against data with differences in noise under the influence of noises or at acquisition timings.

Keywords: smartphones, voice, ear, Multimodal biometric authentication

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

に伴い、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。その上、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情報も保存されている可能性がある。したがって端末を悪意のある他人に利用されると、端末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

¹ 電気通信大学
The University of Electro-Communications

² 静岡大学
Shizuoka University

a) goma.a@uec.ac.jp

フォンを不正利用から守るために、モバイル端末での高精度な個人認証が必要である。

現在スマートフォン上の個人認証には、パスワードやPIN、パターンなどの知識に基づく認証が主に用いられている。しかしパスワードやPINは、簡単なものを利用している場合も多く、第3者に推測される可能性がある。また電車内等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等の認証には盗難等の危険が存在する。そこで近年注目されている認証方法の一つにバイオメトリック認証がある。バイオメトリック認証は紛失、盗難等の恐れが低いと、従来の知識に基づく認証と比べて利便性や安全性が高い。

現在、スマートフォン上の主流なバイオメトリック認証は、指紋認証や顔認証である。多くのユーザーは端末に対してロックをかけているが、その端末に一度認証してしまうと、その後のユーザーが正規のユーザーであるかを検証する事が少ない。また米国において、スマートフォンユーザーの34%が端末のロックしていないと推定されている[2]。よって正規のユーザーではない人物に端末を操作され、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継続的に認証を行うことが考えられる。

例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、通話を行うたびに継続的に認証を行うことが可能である。しかし音声を単体で用いて認証を行う場合、周囲の雑音等により、認証精度に悪影響を与えてしまう。またスマートフォンは家の中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りに利用するものであり、スマートフォンでの音声認証を考える上で雑音の影響は免れない。加えて、通話時に行う音声認証では録音された音声を利用して、他人になりすますことで認証が行われてしまう場合もある。

そこで本研究では他の生体認証と組み合わせて認証を行う、マルチモーダルバイオメトリック認証を考える。音声だけでなく他の生体情報と組み合わせて認証することで、精度が向上する可能性があり、なおかつなりすまし対策になりうる。ところが、一般的に複数の生体情報を組み合わせて認証を行うことで、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ーに負担をかけることなく取得できる生体情報として耳介が挙げられる。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、ユーザーに負担をかけることなくシームレスかつ高精度な認証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み合わせと比較して利便性も向上すると考えられる。

2. 先行研究

2.1 静電容量画像を用いた個人認証

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認証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認証を行うという認証方法である。この認証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認証に比べて特別な認証機器が必要ないという利点が存在する。指紋認証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と、掌などの指紋よりも比較的大きな体の一部を取得することができる点で補うことができる[3]。

Holzら[3]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、それぞれで認証を行った。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。耳のみを用いた認証では、被験者12人に対し、本人拒否率が7.8%であった。

Guo[4]らは、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本の指を用いて図3.2のように認証を行った。取得した画像から550個の特徴を抽出し、その中から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認証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調査を行った。20人の被験者に対して、FRR4.5%のとき、認証精度は94.5%であった。

Tartz[5]らは、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認証を行った。取得した画像から静電容量値の差を利用し、各指の部分を取り出し、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[6]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、図3.4のような耳と指の8bit、 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、耳の長さや幅、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 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

認証が 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認証

岩野ら [7] 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。音声データは 4 桁連続数字を静寂な室内で収録し、16kHz, 16bit で標準化、量子化した。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複数の白色雑音を付加させた。また、耳介画像は右耳正面からデジタルカメラで解像度 720 × 540 の画像を撮影した。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, 8bit 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、MFCC12 次元、MFCC12 次元、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 30dB の場合が最も高精度であった。

また宮崎ら [8] 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. 提案手法

先行研究 ([3], [4], [5], [6]) 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認証方法が研究されているが、スマートフォンでの認証においてそれらを他のモダリティと組み合わせた認証方法は行われていない。また先行研究 ([7], [8]) では、音声と耳介のマルチモーダルバイオメトリック認証を行っているが、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認証と、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認証を考える。

本研究では、継続的に認証を行うことを想定し、主なスマートフォンの利用用途である通話に着目する。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、通話の際に認証することで、継続的な認証に利用可能である。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。ここで通話する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。よってスマートフォン上で用いる、音声と静電容量式タッチスクリーンから取得した耳介画像のマルチモーダルバイオメトリック認証を提案する。音声と静電容量画像を組み合わせて認証することで、精度と利便性の向上が目指せる。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、雑音の影響下でも認証が可能であり、かつ経年変化の少ない認証方法の実現を目指す。統合手法としては、マルチモーダルバイオメトリック認証の先行研究 ([7], [8]) で利用されていた重み付け和によるスコア統合のほか、サポートベクターマシン (SVM) などの機械学習によるスコア統合を用いて評価を行う。

4. 実験

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認証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認証との認証精度の比較を行った。また音声に関しては雑音の程度を変化させ、雑音影響下における認証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。音声は SONY のエレクレットコンデンサーマイクロホン [9] を用いて、サンプリング周波数 44.1kHz, 16bit で ATR 文章 [10] の 50 文を 1 文ずつ録音した。

同時に 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、[11] で公開されていたソースコードを利用した。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。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 個をテストデータとし、残りの 40 個を学習データとし、5-fold cross-validation を行った。また音声、耳介の各スコアは、z スコア正規化を行ったあとに min-max 正規化を行った。

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[12] 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 比で 30dB の雑音を付加した。雑音は、電子協騒音データベース [13] 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 比で 30dB の雑音を加えたの 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 MFCC12 次元、MFCC12 次元、MFCC12 次元、対数パワー 1 次元、対数パワー 1 次元、対数パワー 1 次元計 39 次元の特徴を抽

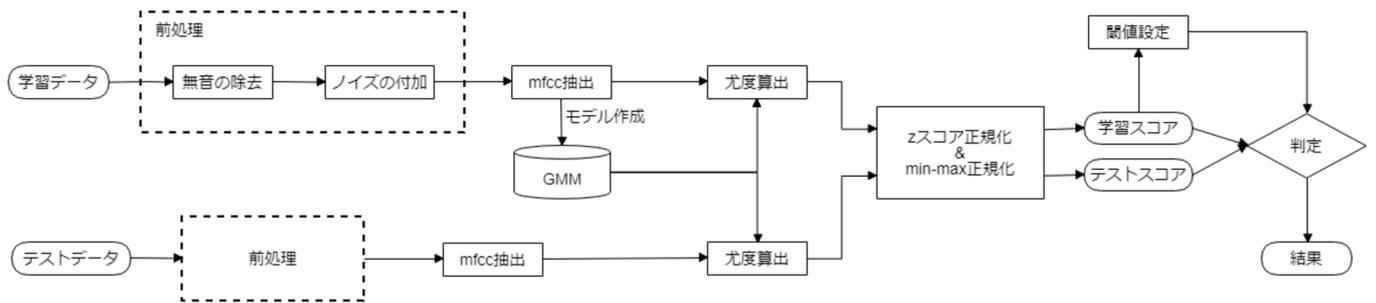


図 1 音声の識別器の構成図
Fig. 1 Diagram of Voice.

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 との尤度を算出した．特徴抽出からスコア算出までは SPTK-3.10[14] 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．取得した耳介のデータは，1 フレームの解像度は原則 24×15 である．静電容量値が 0 以上より小さい場合は 0 に，255 より大きい場合は 255 に変更し，全静電容量値を 0-255 の範囲に補正した．また解像度が 24×15 でない場合は，足りない部分を 0 で埋めた．またデータの取得では，静電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]．その後 1 データ内の全フレームをマージし，平均フレームを作成した．前処理の最後にバイキュービック法を用いて，平均フレームの解像度を 384×240 とした．最終的に 1 人につき 50 枚の画像を生成した．

画像特徴量には SURF[15] を用いた．1 つのテストデータに対して人ごとの学習データとの距離を算出し，そ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，同じ画像同士の距離算出は除く．

4.4 スコア統合

スコア統合の概略図を図 3 に示す．本研究では，先行研究 ([7], [8]) で用いられていた重み付け和に加えて，機械学

表 1 実験 A の認証精度の結果

Table 1 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
SVM	99.50	98.49	94.57	96.41	1.40
KNN	99.42	99.28	92.57	95.68	3.71
重み付け和	99.29	95.85	94.43	94.96	2.23

習でよく用いられているサポートベクターマシン（SVM）と k 近傍法（KNN）を用いて認証精度の比較を行った．音声，耳介の各スコアを連結し，2 次元ベクトルで表したものを機械学習の入力として用いた．

5. 結果

以下に音声に対して雑音環境を変化させて行った実験結果を示す．雑音を付加していない音声を用いた実験を実験 A，雑音 1，雑音 2 をそれぞれ付加した音声を用いた実験を実験 B，実験 C とする．なお，耳介はすべての実験において共通である．

5.1 実験 A の実験結果

表 1 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認証とそれらを組み合わせた認証の結果を示す．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 Accept 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 1 より，Accuracy，Precision，Recall，F1-score，EER の全てにおいて，単体での認証よりも音声，耳介を組み合わせ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

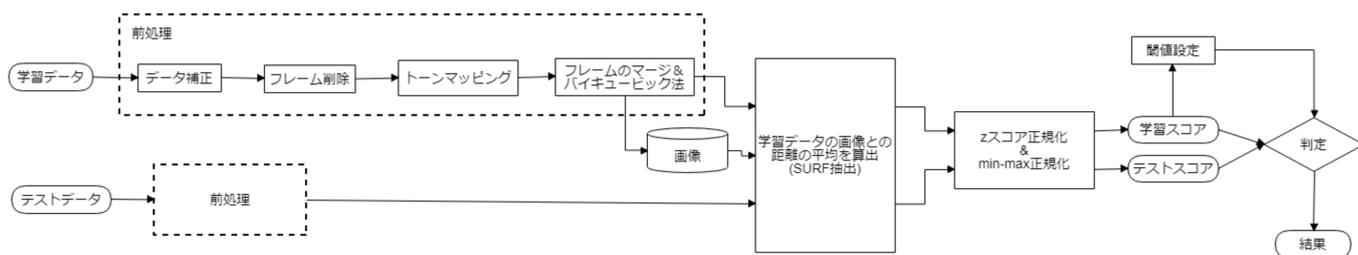


図 2 耳介の識別器の構成図

Fig. 2 Diagram of Ear.

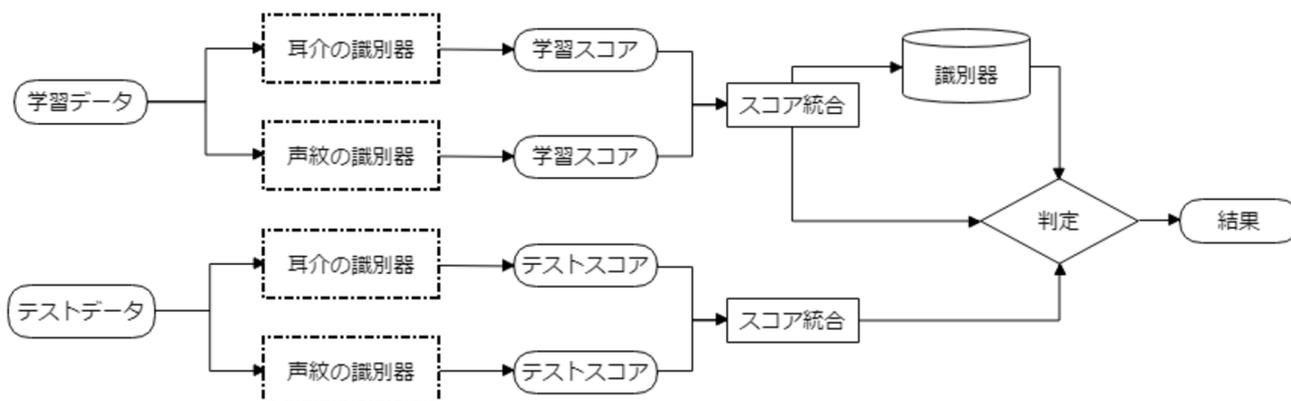


図 3 スコア統合の概略図

Fig. 3 Schematic Diagram of Score Fusion.

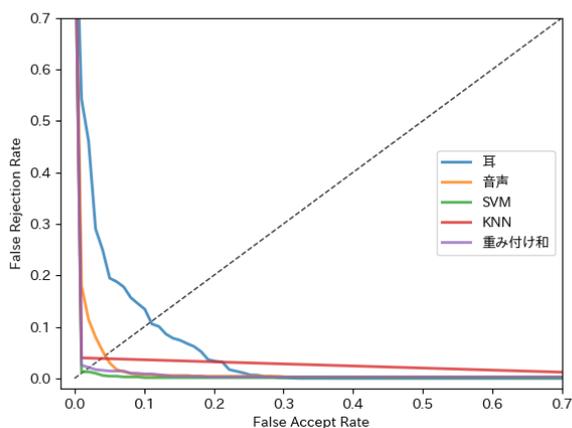


図 4 実験 A の ROC 曲線

Fig. 4 ROC Curve of Experiment A.

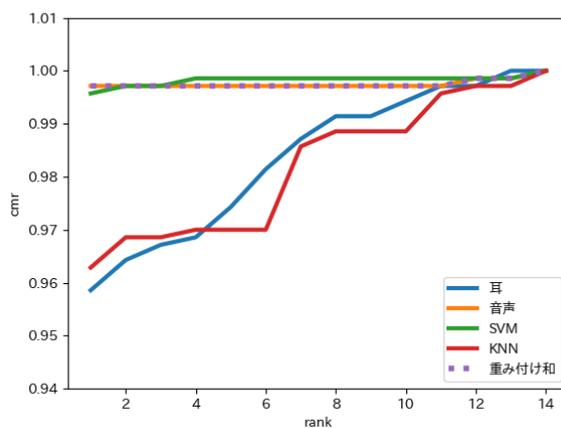


図 5 実験 A の CMC 曲線

Fig. 5 CMC Curve of Experiment A.

合手法は、耳介、音声のそれぞれを用いた認証と比べると 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認証と比べるとおよそ 9.90% , 2.94% 向上した。さらに図 4 からは単体での認証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、SVM、重み付け和の識別率が高いことが確認できた。

5.2 実験 B の結果

表 2 に、雑音 1 を付加した音声と耳介を組み合わせた認証の結果を示す。また図 6、図 7 に ROC 曲線、CMC 曲線を示す。

表 2 より、Accuracy、Precision、Recall、F1-score、EER の全てにおいて、単体での認証よりも音声、耳介を組み合わせ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

表 2 実験 B の認証精度の結果
Table 2 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
SVM	99.53	98.66	94.86	96.63	1.33
KNN	99.42	98.98	92.86	95.70	3.36
重み付け和	99.29	95.85	94.43	94.96	1.61

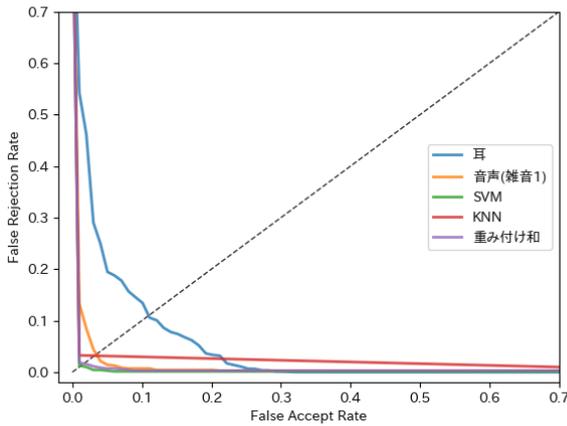


図 6 実験 B の ROC 曲線
Fig. 6 ROC Curve of Experiment B.

表 3 実験 C の認証精度の結果
Table 3 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
SVM	99.58	99.71	94.43	96.94	1.01
KNN	99.54	99.57	94.00	96.58	3.36
重み付け和	99.45	97.27	95.14	96.09	1.51

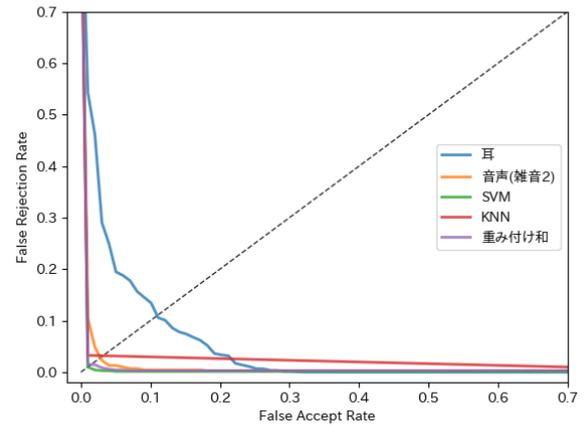


図 8 実験 C の ROC 曲線
Fig. 8 ROC Curve of Experiment C.

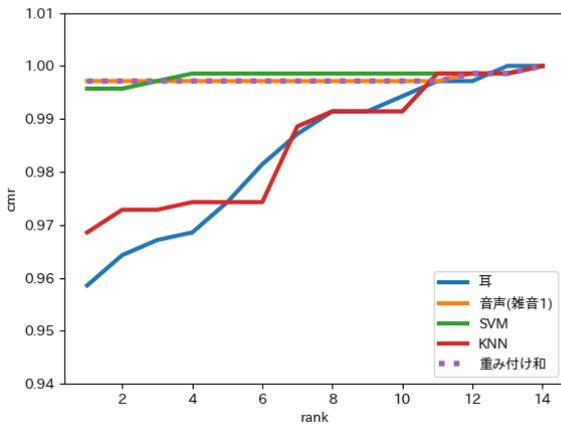


図 7 実験 B の CMC 曲線
Fig. 7 CMC Curve of Experiment B.

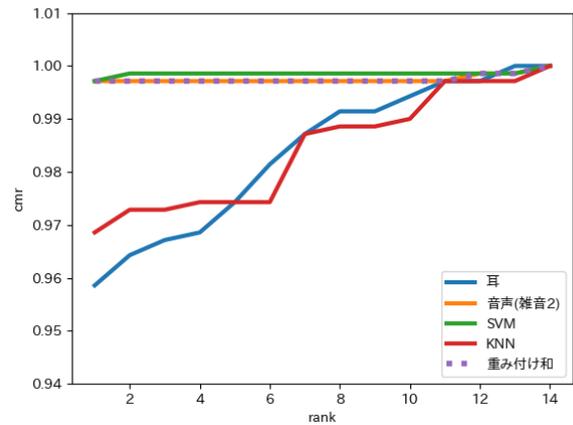


図 9 実験 C の CMC 曲線
Fig. 9 CMC Curve of Experiment C.

合手法は、耳介、音声のそれぞれを用いた認証と比べると 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、音声のそれぞれを用いた認証と、比べるとおよそ 9.97% , 2.08% 向上した。さらに図 6 からは、単体での認証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、SVM、重み付け和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 3 に、雑音 2 を付加した音声と耳介を組み合わせた認

証の結果を示す。また図 8、図 9 に ROC 曲線、CMC 曲線を示す。

表 3 より、Accuracy、Precision、Recall、F1-score、EER の全てにおいて、単体での認証よりも音声、耳介を組み合わせ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認証と比べると 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認証と比べると、およそ 10.29% , 2.04% 向上した。さらに図 4 から単体での認証よりも統合手法のほ

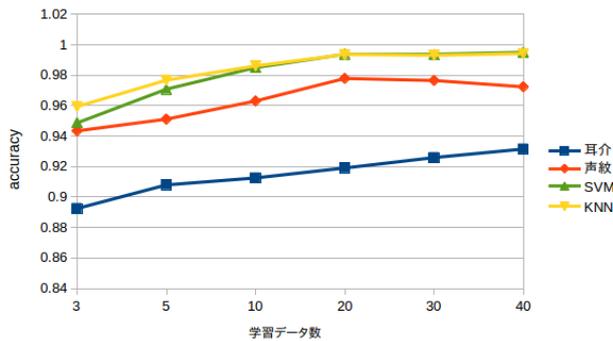


図 10 学習データ数と精度の変化

Fig. 10 The Change of the Number of Training data

うが精度が向上していることが確認でき、図 5 では、音声、SVM、重み付け和の識別率が高いことが確認できた。

6. 考察

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認証である。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、経年変化の 3 つの影響についてそれぞれ考察する。

6.1 学習データ数の変化による精度への影響

本研究では学習データ数を 40 個としたが、スマートフォンでの利用を考えると、生体情報の登録回数が多いほどユーザーへの負担が大きい。よって学習データ数が少ないほど、ユーザーへの負担が少なくより実用的な認証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、各認証での精度の変化を比較した。音声、耳介のそれぞれを単体で用いた認証と、音声と耳介を組み合わせた認証 (SVM) の Accuracy の変化を図 10 に示す。

図 10 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認証することで、音声、耳介の単体での認証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関わらず、単体での認証よりも高精度な認証を行うことが可能である。ここでスマートフォンは日常的に利用するものであるため、最初に登録された生体情報量が少なくても、繰り返しスマートフォンを利用することで、学習データとなりうる生体情報を蓄積させることが可能である。よってスマートフォンを普段利用しているうちに学習データ数を増やすことが可能であり、スマートフォンの通常利用によって、ユーザーに負担をかけることなく、高精度な認証を行うことが可能である。

6.2 雑音の影響

本実験では、雑音のない音声や SN 比で 30dB の雑音を

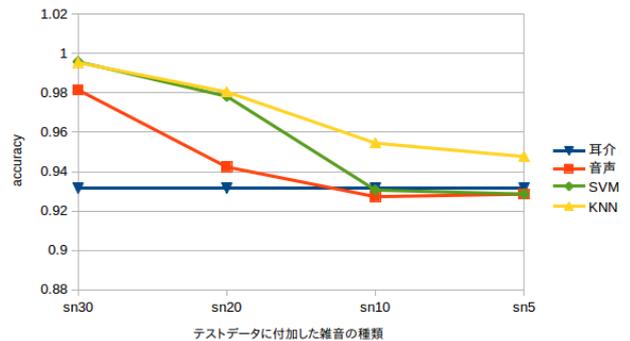


図 11 雑音の大きさによる認証精度への影響 1

Fig. 11 Influence on the authentication accuracy by the noise level part 1

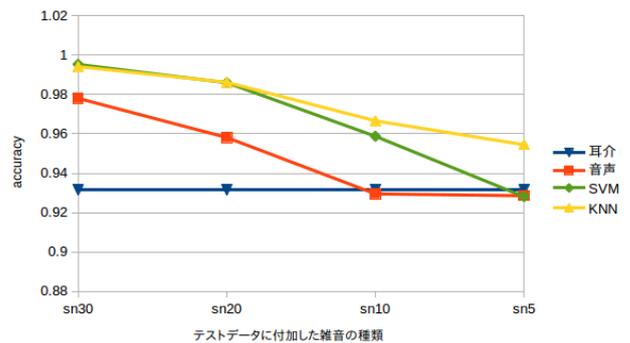


図 12 雑音の大きさによる認証精度への影響 2

Fig. 12 Influence on the authentication accuracy by the noise level part 2

付加させた音声という、雑音が小さい音声を学習データとして用いた。しかしスマートフォンはさまざまな環境下で利用するため、本実験のような雑音が少ない環境だけでなく、雑音が強い環境で認証を行場合もある。そこで、雑音の大きさの変化による認証精度の比較を行った。

まず、雑音を加えていない音声を学習データとした場合の認証精度の結果を示す (図 11)。テストデータには、雑音を加えていない音声、SN 比で 5, 10, 20, 30dB の雑音を加えた音声の 5 種類を用いた。なお音声の雑音の変化にかかわらず、耳介はすべて共通である。

次に、SN 比で 30dB の雑音を付加した音声を学習データとした場合の認証精度の結果を示す (図 12)。テストデータには、SN 比で 5, 10, 20, 30dB の雑音を加えた音声の 4 種類を用いた。

図 11, 図 12 より、単体での認証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、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認証は、様々な場所での認証において、単体での認証よりも雑音の影響を低減

表 4 データ 3 の Precision の変化
Table 4 Change of Precision in Data 3.

単位：%	実験 1	実験 2	実験 3
耳介	35.47	35.47	35.47
音声	95.56	95.04	98.00
SVM	93.18	100.00	100.00

表 5 テストの雑音を強くした場合の Precision の変化
Table 5 Change of Precision in Data 3 When Testdata More Noisy Than Traindata.

単位：%	データ 2	データ 3
耳介	100.00	35.47
音声	100.00	56.67
SVM	100.00	80.00

することが可能である。

6.3 時期差のあるデータへの性能

経年変化の影響を確認するために、およそ 7 か月後に被験者 14 人のうちの 2 人に対して再度データの取得を行った。取得時期が違う 2 人分の 2 種類のデータに対して、最初に取得した 14 人分のデータをデータ 1、データ 1 のうち、再度データを取得した 2 人分のデータをデータ 2、6 ヶ月後に追加で取得した 2 人分のデータをデータ 3 とする。表 4 にデータ 1 で学習し、データ 3 でテストした場合の結果を示す。また、雑音を付加していない音声と耳介を組み合わせた認証を実験 1、雑音 1 を付加した音声と耳介を組み合わせた認証を実験 2、雑音 2 を付加した音声と耳介を組み合わせた認証を実験 3 とする。

実験 1 では精度の向上は見られなかったが、実験 2、実験 3 では提案手法において精度の向上が確認できた。

次に実験 3 において、データ 1 に付加した雑音よりも大きい雑音をデータ 2、データ 3 に付加し、テストを行った場合の結果を示す(表 5)。

データ 2 の結果はすべて 100% であるため単体での認証と提案手法に差はないが、データ 3 では提案手法が耳介認証と比べておよそ 44.53%、音声認証と比べて 23.33% 向上した。以上より、この提案手法は取得時期に差のあるデータでも認証が可能である。また雑音影響下においても、単体での認証より認証精度の改善が見られた。すなわち提案手法はスマートフォンで利用するにあたって、経年変化の影響を低減することが可能である。

7. まとめ

本稿では、スマートフォン上で利用する、ユーザーへの負担が小さくかつ高精度な個人認証として、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認証を提案した。またその性能を評価し、音声、耳介を単体で用いた認証と比較し、

提案手法の有効性を確認した。雑音影響下や、取得時期に差のあるデータに対しても有効であることも確認できた。

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挙げられる。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、および認証精度への影響の評価も挙げられる。

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: <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/html/nc262110.html>
- [2] Upal Mahbub et.al: *Active user authentication for smartphones: A challenge data set and benchmark Results*, BTAS2016 (2016).
- [3] Christian Holz et.al: *Biometric User Identification on Mobile Devices Using the Capacitive Touchscreen to Scan Body Parts*, CHI2015 (2015).
- [4] Anhong Guo et.al: *CapAuth: Identifying and Differentiating User Handprints on Commodity Capacitive Touchscreens*, ITS2015 (2015).
- [5] Robert Tartz et.al: *Hand Biometrics Using Capacitive Touchscreens*, UIST2015 (2015)
- [6] Mohamed Azaad Rilvan et.al: *User authentication and identification on smartphones by incorporating capacitive touchscreen*, IPCCC2016 (2016).
- [7] 岩野 公司: 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 年春季講演論文集 Vol. No3-3-3 pp.109-110 (2003).
- [8] 宮崎 太郎: 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 Vol. No2-4-7 pp.99-100, 200
- [9] ECM-PCV80U 主な仕様 — マイクロホン — ソニー: <http://www.sony.jp/microphone/products/ECM-PCV80U/spec.html>
- [10] ATR503 文 - 音素バランス文・語 - 音声資源コンソーシアム: <http://research.nii.ac.jp/src/ATR503.html>
- [11] RainCheck:<http://isaaczinda.com/raincheck/index.html>
- [12] Audacity:<https://www.audacityteam.org/>
- [13] 電子協騒音データベース: <http://shachi.org/resources/4313?ln=jpn>
- [14] Speech Signal Processing Toolkit (SPTK), <https://sourceforge.net/>
- [15] SURF
- [16] トーンマッピング - *t-pot*, http://t-pot.com/program/123_ToneMapping/index.html
- [17] 山内 拓也: 高ダイナミックレンジ画像のための注視領域を用いたトーンマッピング手法の評価, 日本写真学会誌 75 巻 1 号 p.87-96, (2012).
- [18] 奥田正浩: *HDR 画像 ~ 色空間から符号化まで ~*, https://www.jstage.jst.go.jp/article/itej/64/3/64.3.299/_pdf
- [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: <http://kikakurui.com/x8/X8101-1-2010-01.html>