スマートフォンにおける短い発話時間での音声と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．そこで，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オメトリック認証が利用されている．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認証を提案する．通話中に音声，耳介データを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素早く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精度をテストデータの量を変化させて比較，評価し，認証に必要な発話時間を 30%程度に削減できた．提案手法では，短い発話時間を用いた場合でも高精度に認証可能であることを確認した．

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オメトリック認証

Personal Authentication on Smartphone Using Voice and Ear in a Short Utterance Time

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

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 speedy and highly accurate authentication without imposing a burden on the user by using natural actions in daily life . Compared and evaluated the accuracy of authentication of the proposed method and voice and auricle by changing the amount of test data, the speech time required for authentication was reduced to about 30It was confirmed that the proposed method can authenticate with high accuracy even when using a shorter utterance time.

Keywords: smartphones, voice, ear, multimodal biometric authentication

1. はじめに

近年，スマートフォンが広く普及し，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が増加している．さらに，スマートフォン内には他人の連絡先などの個人情報や 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．したがってスマートフォン内の情報を他人に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，高精度な個人認証が必要である．

[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

現在スマートフォン上での個人認証には，主にパスワードや PIN，パターンなどの知識に基づく認証が用いられている．しかしこれらの認証では，忘れないために簡単な文字列やパターンを設定する場合も多く，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解除の動作を他人に盗み見られ，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ため，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，スマートフォン上での主流なバイオメトリック認証は，指紋認証と顔認証である．これらの認証は主にスマートフォンのログイン時に利用される．多くのユーザーはスマートフォンに対してロックを施すが，ログイン時に一度認証を行うと，その後のユーザーが正規のユーザーであるかということ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%がロックをしないと推定されている [1]．よって他人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を利用して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．しかし音声のみを用いて認証を行う場合，周囲の雑音等により，認証精度に悪影響が与えられてしまう．またスマートフォンは様々な場所で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．

さらにもう一つの課題として，認証に用いる発話時間の長さが挙げられる．発話時間が長いほど，認証に用いるデータ量が多くなる．しかし発話時間の増加に伴い，ユーザーの心理的，物理的ストレがス増加する恐れがある．さらにスマートフォンを他人に不正利用されている場合，短い時間で認証できるほど，より早く他人の利用を防ぐことが可能である．よって認証時間が短いほど，セキュリティ面でも良い効果が期待される．したがって，より短い認証時間で高精度に認証が可能な認証方法であるほど，ユーザーの負担が軽減され，かつセキュリティが向上した認証方法であると考える．

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により，認証精度を向上させ，認証に用いる発話時間が短い場合でも認証精度の低下を防ぎ，高精度に認証を行うことが可能である．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素早くかつ高精度に認証を行うことが可能となり、ユーザブルセキュリティを高めることができる．

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る認証方法である．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ため，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 [2].

Holz ら [2] 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．

Guo[3] 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も調査を行った．

Tartz[4] らは，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，各指の部分を切り出し，指ごとに特徴量を抽出した．

Rilvan ら [5] 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

表 1 1 テストデータあたりの量
Table 1

	100%	50%	30%	10%
音声（秒数 [s]）	3.56	1.78	1.07	0.36
音声（フレーム数 [枚]）	1425	713	428	143
耳介（フレーム数 [枚]）	133	67	40	14

ををを 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．

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 [6] 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 で標本化，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．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，Δ MFCC12 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．また宮崎ら [7] 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. 予備実験

まずテストデータの量を変更することで，通話時間が短くなった場合に起こる認証精度への影響を調査した．前処理後のテストデータの量をすべて一律に元の量から 50%，30%，10%に削減した．前処理後の 1 テストデータあたりの平均量を**表 1** に示す．

3.1 使用データ

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 [8] を用いて，サンプリング周波数 44.1kHz，16bit で ATR 文章 [9] 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，[10] 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．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た．また残りの 40 個のうち，3 個を学習データ 1，30 個を学習データ 2 とし，5-fold cross-validation を行った．また音声，耳介の各スコアに対して z スコア正規化を行ったあとに min-max 正規化を行った．

3.2 音声の識別器

図 1 に音声の識別器の構成を示す．

まず始めに，Audacity[11] 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 [12] の人混みを用いた．そして各種類ごとに特徴量として MFCC12 次元，Δ MFCC12 次元，Δ Δ MFCC12 次元，対数パワー 1 次元，Δ 対数パワー 1 次元，Δ Δ 対数パワー 1 次元の計 39 次元の特徴量を抽出した．学習データ 1 を利用し，被験者ごとに混合ガウスモデル（GMM）を作成し，学習データ 2，テストデータと 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[13] を利用した．また，各モデルとの学習データ 2，テストデータの尤度を，それぞれ学習，テストスコアとした．学習データ 2 のスコアで EER 付近に閾値を設定し，閾値を超えるスコアを本人，下回るスコアを他人と判定した．

3.3 耳介の識別器

図 2 に耳介の識別器の構成を示す．

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の 1 フレームの大きさは，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 から 255 の範囲に補正した．また大きさが 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，1 データ内の各フレームにトーンマッピング [14]，[15] 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 [16]．前処理の最後にバイキュービック法を用いて，

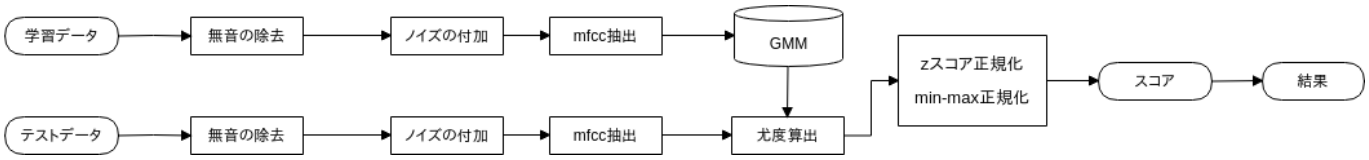

図 1 音声の識別器
Fig. 1 Diagram of Voice.

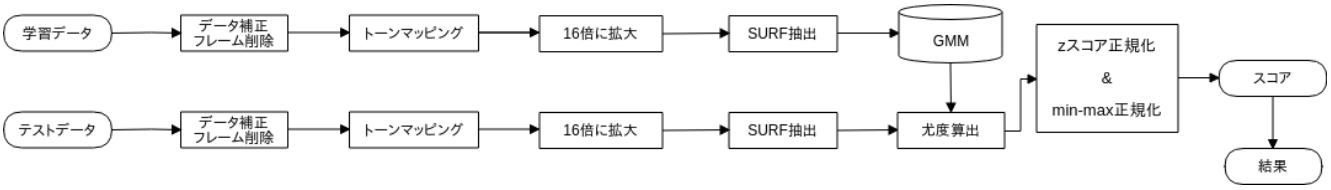

図 2 耳介の識別器
Fig. 2 Diagram of Ear.

表 2 予備実験の結果（EER）
Table 2

単位（%）	100%	50%	30%	10%
音声	7.73	7.23	8.94	14.12
耳介	9.82	10.05	10.08	10.95

各フレームを 16 倍に拡大した．

画像特徴量には SURF[17] を用いた．学習データ 1 の各フレームから surf を抽出し，GMM を作成した．また 3.2 と同様に，各モデルとの学習データ 2，テストデータの尤度を，それぞれ学習，テストスコアとした．学習データ 2 のスコアで EER 付近に閾値を設定し，閾値を超えるスコアを本人，下回るスコアを他人と判定した．

3.4 結果

表 2 に結果を示す．

音声，耳介のどちらにおいても，テストデータの量を削減すると 100%の量のテストデータと比較して，認証精度は低下した．特に 10%にまで削減すると 100%と比較して，耳介は 1.13%，音声は 6.39%低下した．耳介に比べて音声の精度が大幅に低下していることが分かる．よってテストデータの量が減少した場合でも高精度に認証を行うためには，音声の認証精度を改善する必要があると考える．

4. 提案手法

本研究では，スマートフォンにおいて認証時間が短い場合でも認証精度の低下を緩和可能な認証方法を目指す．認証精度の低下を緩和するために，まず音声以外の生体情報と組み合わせて認証を行うことでより高精度な認証方法とし，次に音声そのものの認証精度を向上させる．

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する耳介データと通話時に取得する音声データを組み合わせたマルチモーダルバイオメトリック認証により，短い認証時間においても高精度であり，より負担の少ない

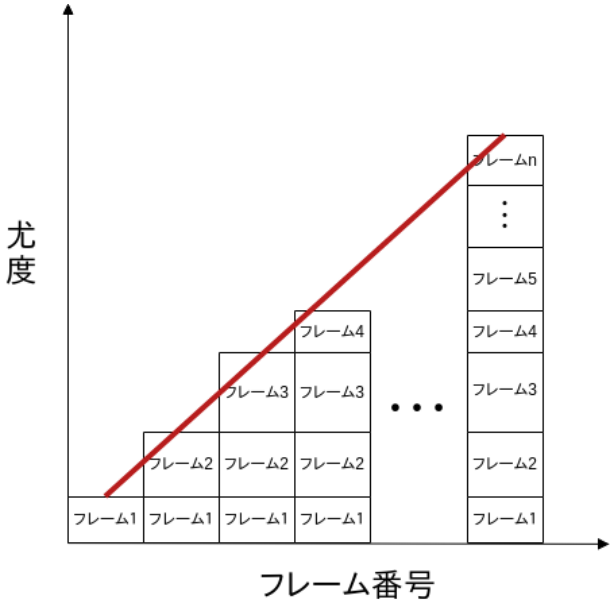

図 3 音声の傾きイメージ
Fig. 3 Image of Likelihood Slope

認証方法を提案する．さらに，音声において，3.2 のように GMM から求めた尤度に加えて，尤度の傾きを新たな特徴として利用することで，より高精度に認証を行うことが可能である．尤度の傾きは，各フレームの尤度を足し合わせることによって求めることが可能である．**図 3** に音声の傾きのイメージ図を示す．例えば 1 フレームと 2 フレームの尤度を足した尤度を 2 フレーム目の尤度とし，同様に 1 から n フレームまでの尤度を足したものを n フレーム目の尤度とすることで，図 3 のようになり，尤度の増加量が求まり，傾きが得られる．一般的に本人のモデルに対する尤度は大きく，他人のモデルに対する尤度は小さい．よって本人と他人それぞれのモデルに対する尤度の積み重ねを比較すると，本人のモデルに対する尤度が大きいため、増加量も大きくなり，傾きが大きくなる．よって傾きを利用することで，発話時間が短くても高精度に認証が可能である．

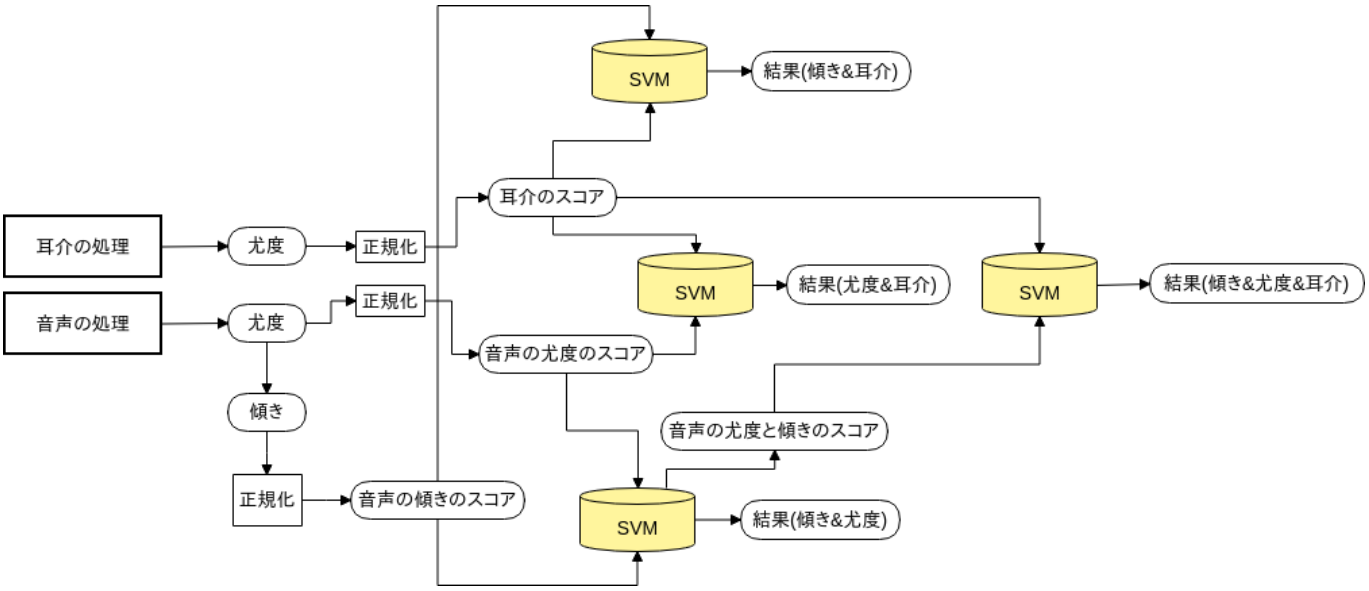

図 4 スコア統合
Fig. 4 score fusion

また音声の尤度と傾き，耳介のそれぞれを組み合わせ，サポートベクターマシン（SVM）を用いてスコア統合を行う．音声，耳介の識別器から得られたスコアをベクトルで表し，そのベクトル空間において SVM で認証を行う．

5. 実験

スマートフォンでの継続的な認証において，認証時間が短い場合でも高精度に認証ができることを示すために，1 データごとにデータ量を変化させ，音声の被験者全員のデータ量を揃えて実験を行った．音声では，100 から 700 フレーム分まで 100 フレームごとに計 7 種類の長さにのデータを利用し，1 データの総フレーム数に対する割合をそれぞれ求めた．耳介では，音声で求めた割合から対応する耳介 1 データ分のフレーム数を利用した．例えば，被験者 A のテストデータ 1 の音声データが全体で 1000 フレームであった場合，100 フレームは全体の 10%であり，200 フレームは 20%である．ここで被験者 A のテストデータ 1 の耳介データが全体で 100 フレームであった場合，音声データの 100 フレーム分に相当する耳介データは 10 フレームである．同様に音声で 200 フレームを用いる場合は，耳介は 20 フレーム用いた．このように，テストデータの量を音声を基準にして調整し，各フレーム数ごとに認証精度の比較を行った．

5.1 使用データ

データは 3.1 と同様のデータを使用した．

5.2 音声の識別器

3.2 と同様の処理を行った．加えて，4 で示した傾きについては，1 フレーム目から最後のフレームまでの各フレームの尤度を足し合わせ，傾きを求めた．傾きも尤度と同様の処理により，スコア算出を行った．

5.3 耳介の識別器

3.3 と同様の処理を行った．

5.4 スコア統合

音声，耳介の各スコアの統合には，機械学習でよく用いられているサポートベクターマシン（SVM）を用いた．5.2，5.3 で得たスコアのうち，学習データ 2 から得られるたスコアを SVM の学習に用いた．音声，耳介の各スコアを 2 次元ベクトルで表し，機械学習への入力として用いた．スコア統合は，音声の尤度と傾き，音声の尤度と耳介，音声の傾きと耳介，音声の尤度と傾きのスコア統合と耳介の 4 種類を行い，尤度，傾き，耳介のそれぞれとの認証精度と比較した．**図 4** にスコア統合の概略図を示す．

5.5 結果

表 3 にテストデータの量をフレーム数で削減した実験の結果を示す．

表 3 より，全体的にフレーム数が多くなるほど精度が向上した．音声と耳介それぞれでは，音声の方が高精度であった．また，音声の傾きと尤度それぞれにおいては，大きな差はなかった．しかし，音声の傾きと尤度をスコア統合することで，認証精度が向上し，700 フレームの場合，尤度のみ，傾きと比較して，それぞれ 0.11%，0.1%向上した．さらに音声と耳介をスコア統合することで，音声，耳介のそれぞれと比較して認証精度が向上した．100 フレームの時点では音声の傾きと耳介のスコア統合が最も高精度であり，音声の傾きと尤度のスコア統合と比較して 3.89%，耳

表 3　テストデータを削減した結果（EER）
Table 3　Result of reducing test data(EER).

フレーム数	100	200	300	400	500	600	700
音声（傾き）	17.57	11.45	9.86	8.88	7.88	7.19	7.01
音声（尤度）	17.36	11.60	9.87	8.84	8.01	7.02	7.02
音声（傾き&尤度）	17.18	11.37	9.66	8.43	7.78	6.93	6.91
耳介	18.31	13.78	12.45	11.26	10.41	10.42	10.40
SVM（傾き&耳介）	13.29	10.05	7.62	6.11	5.43	4.78	4.54
SVM（尤度&耳介）	14.67	11.38	7.90	6.56	5.62	4.78	4.56
SVM（傾き&尤度&耳介）	14.04	11.18	7.80	6.43	3.98	3.34	3.67

介と比較して 5.02%向上した．また特に 500 フレーム以降では，傾きと尤度のスコア統合と耳介のスコア統合が高精度であり，700 フレーム時点での精度を比較すると，音声の傾きと尤度のスコア統合と比較して 3.28%，耳介と比較して 6.73%向上した．

6. 考察

3.4 と 5.5 より，認証に必要な発話時間および傾きについて考察する．

6.1 通話時間と認証精度の比較

3.4 と 5.5 を比較する．表 1 より，100%の音声のテストデータは平均で 3.56 秒であり，1425 フレームである．よって 3.4 の 50%，30%，10%とは，5.5 において 700，400，100 フレーム付近に相当する．耳介においては，3.4 の 50%で 10.05%，5.5 の 700 フレームで 10.40%と同等の精度が確認できた．音声においても同様である．

音声の傾きと尤度のスコア統合と耳介のスコア統合に着目すると，5.5 の 400 フレームで 6.43%であり，3.4 の音声 100%と比較して 1.30%上回っている．よって 5.5 において 300 フレーム分のデータを用いた認証は，音声の 100%のデータ量を用いた認証と同等以上の認証精度であり，データ量をおよそ 28.1%まで削減することが可能である．したがって音声と耳介を組み合わせた認証方法は，音声と耳介のそれぞれのみの認証よりも高精度であることに加えて，短い認証時間で認証時間が長い認証方法と同等以上の精度で認証が可能であると言える．また，音声の傾きと尤度のスコア統合と耳介のスコア統合においては，5.5 の 700 フレームで 3.67%であり，3.4 の音声 100%と比較すると，およそ半分の認証時間で 100%のデータ量の認証精度の 2 倍の認証精度が得られた．

また Accuracy でも比較を行う．**表 4**，**表 5** に，表 2，表 3 に対応する Accuracy の結果を示す．

Accuracy の場合も，EER と同様の傾向が得られた．音声は尤度のみでなく傾きも利用することで，100%と同等の認証精度を 300 フレームで得られた．また耳介と組み合わせることで認証精度も向上し，200 フレームで表 5 の音声

表 4　予備実験の結果（Accuracy）
Table 4　Results of preliminary experiments(Accuracy).

単位（%）	100%	50%	30%	10%
音声	93.70	91.73	90.32	84.24
耳介	92.21	91.74	91.90	91.69

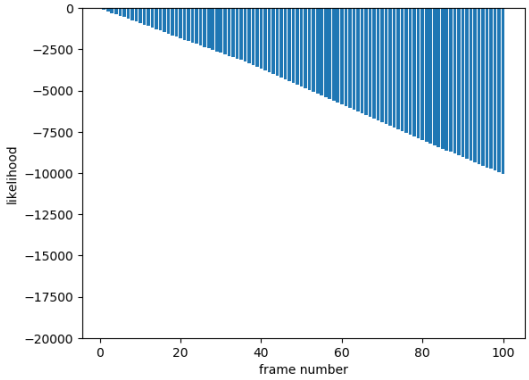

図 5　本人のモデルとの尤度
Fig. 5　Likelihood with His Model.

100%の精度である 93.7%を上回る認証精度が得られた．

以上より，音声と耳介を組み合わせた認証方法は，それら単体のみの認証方法を比較して，認証に必要な発話時間が短い場合でも高精度な認証方法である．さらに，音声の傾きを新たな特徴として追加することで，より高精度な認証が可能である．短い時間で高精度に認証が可能であるため，ユーザーの負担が軽減され，より利用しやすい認証方法である．

6.2 音声の傾き

まず傾きの一例として，同じテストデータで本人のモデルと他人のモデルに対する 100 フレーム分の尤度を**図 5**，**図 6** に示す．図 3 では尤度を正の値で示したが，本実験での尤度は，ツールの都合よりすべて負の値であった．そのため，本研究では傾きが緩やかになるほど傾きが大きくなる．

図 5 は本人のモデルとの尤度，図 6 は他人とのモデルの尤度を 100 フレーム分足し合わせたグラフである．本人のモデル，他人のモデルとの尤度を比較すると，図 5，図 6

表 5　テストデータを削減した結果（Accuracy）
Table 5　Result of reducing test data(Accuracy).

フレーム数	100	200	300	400	500	600	700
音声（傾き&尤度）	85.74	92.94	94.46	95.27	95.48	95.63	95.88
耳介	91.56	92.32	92.30	92.27	92.04	92.04	91.95
SVM（傾き&尤度&耳介）	92.05	96.32	97.23	97.38	97.81	97.87	98.00

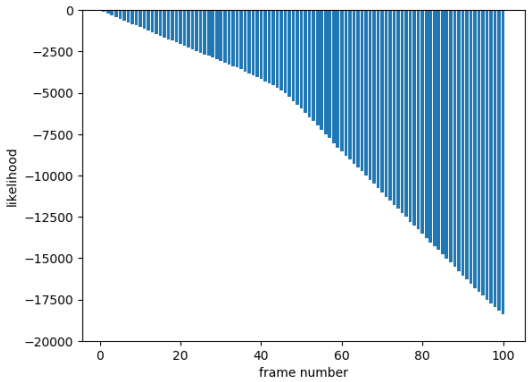

図 6　他人のモデルとの尤度
Fig. 6　Likelihood with Others' Models.

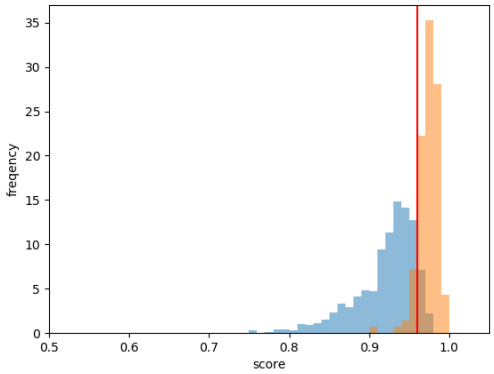

図 7　100%のデータのヒストグラム
Fig. 7　Histogram with 100% Data.

のように本人のモデルとの尤度のほうが大きくなる．よって各フレームの尤度を足し合わせると，本人のモデルとの照合の方が尤度の傾きが大きくなる．ちなみに図 5，図 6 における傾きは，それぞれ-100，-183 である．したがってこのように少ないデータ量であっても，本人同士の照合か，他人との照合かどうかで差が発生するため，特徴として用いることが可能である．

5.5 より，音声の尤度のみではなく，傾きも加えて認証を行うことで，通話中の発話時間が短い場合でも高精度な認証方法であることを確認できた．音声の尤度と傾きそれぞれの認証精度には，大きな差はない．しかし照合結果から，傾きは尤度と比較して recall が，尤度は傾きと比較して precision が高い傾向があった．そこで傾きと尤度を統合することで，総合的に認証精度が向上したと考えられる．

次に，音声のヒストグラムを比較する．**図 7** に 100%のデータ量，**図 8** に 500 フレーム分，**図 9** に 100 フレーム分の傾きヒストグラムの一例を示す．分かりやすくするため，横軸の範囲を調整し，一部を示している．オレンジ色の分布が本人，水色の分布が他人の分布であり，赤い縦線が EER は付近に設定した閾値である．図 9 は図 7 と比較すると，誤って他人が閾値を超えている部分と，本人が閾値を下回っている部分が多いことが分かる．よって 100%のデータよりも認証精度が低下していることがヒストグラムから分かる．しかし図 7 と図 8 を比較すると，誤った部分が同程度である．100%のデータ量，500 フレーム分の傾きを利用した場合の EER は，それぞれ 7.73%，7.88%であり，同等の認証精度であることがヒストグラムからも確認できる．

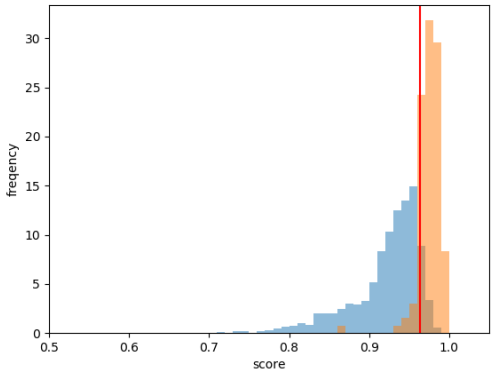

図 8　500 フレーム分のヒストグラム
Fig. 8　Histogram with 500 Frames.

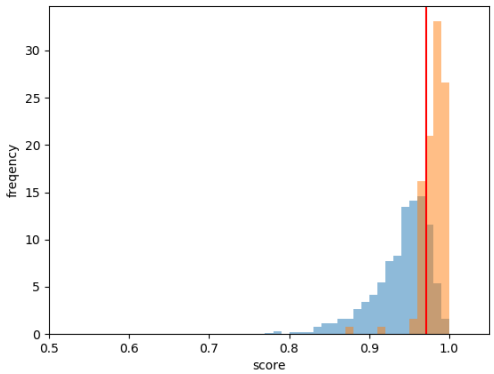

図 9　100 フレーム分のヒストグラム
Fig. 9　Histogram with 100 Frames.

次に，音声の傾きと耳介の散布図を比較する．**図 10** に 100%の学習データ，**図 11** に 500 フレーム分，**図 12** に 100 フレーム分の散布図の一例を示す．本人と他人の分布がわ

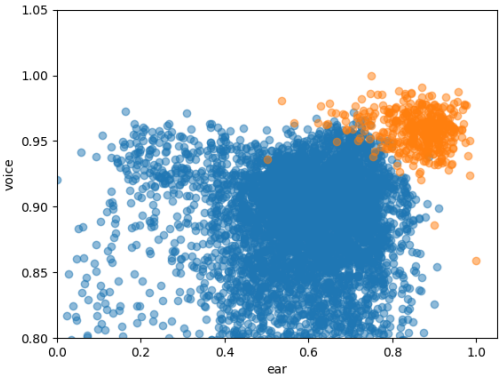

図 10 100%のデータの散布図
Fig. 10 Scatter Plot with 100% Data.

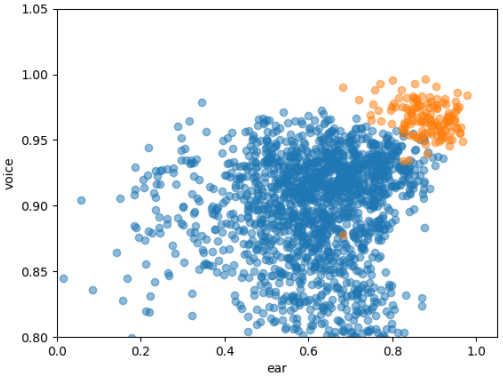

図 11 500 フレーム分の散布図
Fig. 11 Scatter Plot with 500 Frames.

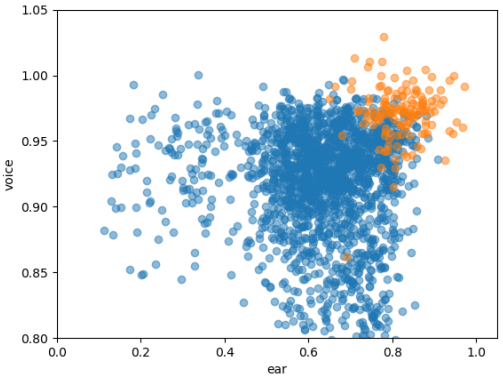

図 12 100 フレーム分の散布図
Fig. 12 Scatter Plot with 100 Frames.

かりやすいように，縦軸と横軸を調整している．オレンジ色の分布が本人，水色の分布が他人の分布である．図 10 と図 12 を比較すると，図 12 の方が他人と本人の分布が重なっており，100 フレームでは認証精度が低下することが分かる．次に図 10 と図 11 の分布を比較すると，概ね類似している．ここで 3.4 の音声と耳介の 100%のデータで SVM を用いてスコア統合した結果は，5.16%であった．よって 500 フレームを利用することで 5.43%と同等の精度が得られたことが，散布図から比較しても確認できる．

7. まとめ

本稿では，スマートフォン上で用いるユーザーへの負担が少なく，かつ高精度な認証方法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．またその性能を比較し，提案手法は，音声，耳介をそれぞれ用いる認証方法より，認証に必要な発話時間を 30%程度に削減でき，素早く高精度に認証可能な認証方法であることを示した．

今後の課題としては，まず正規化の他に外れ値の扱い方を工夫する．また今回音声に傾きを利用したように，音声や耳介で他に利用可能な特徴を検討する．

参考文献

[1] Upal Mahbub et.al.：*Active user authentication for smartphones: A challenge data set and benchmark Results*， BTAS2016（2016）.
[2] Christian Holz et.al.：*Biometric User Identification on Mobile Devices Using the Capacitive Touchscreen to Scan Body Parts*， CHI2015（2015）.
[3] Anhong Guo et.al.：*CapAuth: Identifying and Differentiating User Handprints on Commodity Capacitive Touchscreens*， ITS2015（2015）.
[4] Robert Tartz et.al.：*Hand Biometrics Using Capacitive Touchscreens*， UIST2015（2015）
[5] Mohamed Azard Rilvan et.al.： *User authentication and identification on smartphones by incorporating capacitive touchscreen*， IPCCC2016（2016）.
[6] 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）.
[7] 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
[8] 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
[9] 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
[10] RainCheck:http://isaaczinda.com/raincheck/index.html
[11] Audacity:https://www.audacityteam.org/
[12] 電子協騒音データベース：http://shachi.org/resources/4313?ln=jpn
[13] Speech Signal Processing Toolkit （SPTK），https://sourceforge.net/
[14] トーンマッピング - *t-pot*， http://t-pot.com/program/123_ToneMapping/index.html
[15] 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）.
[16] 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
[17] Herbert Bay et.al.， ：*SURF: Speeded Up Robust Features* ， ECCV （2006）.