携帯端末への話者照合を用いたセキュリティロック*

山室慶太 (法政大学情報科学部), 伊藤克亘 (法政大学情報科学部)

1 まえがき

RFID技術を用いた電子マネーやPIMによる個人情報の管理など，情報技術の発達により，様々な個人情報を携帯電話などの一つの携帯端末で管理・利用することが多くなってきている．これらの端末にはパスワードを認証キーとしたセキュリティが掛かっているがパスワードによる認証では，忘却や漏洩などの恐れがある．そこで生体認証の一つである話者認証に注目した．生体情報の音声を用いる話者認証であれば忘却や漏洩などの危険性を減らすことができる．また，パスワードとの併用によってより強固なセキュリティロックを実装することも考えられる．

話者認証技術は現在までに，雑音に強固な音声認証として文献[1]のような高精度な認証精度を持たせるなどの研究がされている．

しかし，問題点もいくつか残っている．その一つは話者照合システムを構築するための学習データとして多く必要とされる点である．また，人の声は気候など周囲の変化から影響を受けるため声の特徴が変化してしまう問題点もある．

そこで，本論文では電話時の会話音声を学習データとして利用することを提案する．本来モデルの更新のための学習データの収集は周期的に用意する必要があるため話者へ負担を掛けてしまうが，携帯端末での会話音声を学習データとして再利用することで話者に負担を減らすことができる．また，随時録音した音声を利用して繰り返し音響モデルの更新を行う．

2 モデルの更新による話者照合

本論文では話者照合に繰り返し更新を行った音響モデル用いることを想定している．これまでに話者照合の音響モデルの作成方法について文献[4]のような研究がされている．人の音声は時期により特徴が変動してしまい，同じ音響モデルを使い続けていると認識率に影響が出てしまう．この研究ではモデル更新を伴う話者認識方法を用いて複数の時期に録音した10文章を使い周期的にモデルの更新を行った結果，そのモデルの本人棄却率と他人受理率が更新を行わなかったモデルの本人棄却率と他人受理率よりも約8割まで減少し，この手法の有効性を示していた．

3 システムの概要

話者照合を行うシステムは二種類の工程で行われるものを想定している．

一つ目の工程では，音声データを収集している．まず，話者照合に用いる音声データ収集する必要があるためandroid端末上で音声の録音を行う．録音する音声には話者の会話時のものを用いている．録音した音声は一時的にandroid端末内でまとめて保管しておく．この音声データはすべてTCP通信によってサーバー側のPCへ転送される．サーバー側のPCでは受信した音声データの情報や収録日時など，音響モデルに利用する情報を登録し，管理しておく．

二つ目の工程では，話者の照合を行う．まず，収集した音声データを利用して音響モデルの構築を行う．このとき，すでに音響モデルを構築している場合はその際に利用した音声データと新しく登録された音声データを使い，音響モデルの更新を行う．構築した音響モデルはTCP通信によってandroid端末に転送される．この音響モデルとマイク入力された音声データを用いることで話者照合を行う．

4 音響モデルの性能実験

話者照合の識別精度を確認するために音響モデルの性能実験を行う必要がある．累積した様々な音声データにあわせて，音声の変化を表現しやすいようにモデルの構造を複雑化しながら話者モデルを再構築することが理想的である．そのため今回は二つの音響モデルを用意し，それぞれデータを学習させたモデルの性能実験を行い比較した．ひとつは1状態，混合数32の話者単位で構築したGMMである．ふたつめは5状態，混合数4の音節単位で構築したHMMである．それぞれの音響モデルは各話者ごとの音声情報を学習させた話者本人情報を持つモデルと男女20名分の話者情報を学習させた他者の情報を持つモデルを1セットとして考えて構築されている．この音響モデルとテストデータから求めた尤度を閾値と比較することで話者照合を行い，本人棄却率と他人受理率を求めることで性能を評価した．

この実験は通話時などの会話音声を随時録音し，学習データとして蓄積させていくことを想定している．そのため，学習データには日程を変えて録音した音声を複数用意して，それらを時期別に分けた3つのGMMに学習させた．各日程ごとの学習データを使い音響モデルの更新を行うことでモデルの本人棄却率と他人受理率の変化を調査し，更新を行わなかった音響モデルとの比較を行った．

4.1 音声データ

実験に用いた音声データには話者20名に発話してもらった音素バランス文を約一ヶ月間にわたって録音したものを用いた．まず，初期学習用の音声データとして50文の音素バランス文を収録した．その後，複数の日程で録音を行い，それぞれの日程で音響モデルのアッ

* A security lock by speaker verification using a smart phone by Keita Yamamuro. (CIS, Hosei University) et. al.

プデート用学習データを 10 文，テストデータを 20 文収録した．録音したデータのパラメータはサンプリング周波数 16kHz，量子化ビット数 16bit となっている．この音声データは 1 から 12 次元までの MFCC12 次元とその 1 次差分，F0 情報を 1 次元，ΔF0 情報を 1 次元，Δ 対数パワー 1 次元の計 27 次元の音声特徴量をフレーム長 25ms，フレーム周期 10ms で抽出した．音声の録音には adroid 端末 (docomo の HT-03A) の内蔵マイクを使って行った．

4.2 基本周波数（F0）の情報

話者照合に F0 情報を用いることで認識性能の対雑音性が向上することが文献 [2] で報告されている．そのため今回音声特徴量として MFCC のほかに F0 とその差分である ΔF0 の情報を利用した．

F0 情報の抽出には STRAIGHT を用いている．複数の手法を使い雑音の影響下で F0 の情報を求めている文献 [3] において STRAIGHT は高精度な推定を行うことができるという結果が出ている．本研究で扱う音声データには雑音が混ざっているものが多くなる可能性があるため STRAIGHT を使い特徴量を抽出した．また，ΔF0 は F0 情報の前後 10ms ごとの値から最小 2 乗法によって得られる傾きを ΔF0 としている．

4.3 認識性能の評価

話者照合のシステム性能を評価する時，本人が棄却される誤り率（本人棄却率）と他人が受理される誤り率（他人受理率）が用いられる．これら 2 種類の誤り率と判定の閾値との間はトレードオフの関係となっている．

今回この閾値は音響モデルと同じく学習データを用いて毎回更新を行っている．閾値の設定には更新した音響モデルと学習データを用いて一度認識を行い，その結果の尤度から平均した値を用いている．

4.4 実験結果

更新を行った話者単位モデルと未更新の話者単位モデルから本人棄却率と他人受理率を求めた結果を図 1 に示す．この結果は各話者の認識率を別々の日程ごと各 3 回録音したテストデータを用いて算出し，その結果を平均化したものとなっている．

認識率を比較した結果，本人棄却率の平均はそれぞれ未更新モデルが 8.1 %，更新したモデルが 6.0 %となり，モデルの更新を行うことで本人棄却率が未更新のものより約 25.0 %減少した．また他人受理率の平均はそれぞれ未更新モデルが 18.6 %，更新したモデルが 12.6 %となり，モデルの更新を行うことで他人受理率が約 30.0 %減少した．

次に音節単位で構築したモデルの認識率だが，本人棄却率の平均はそれぞれ未更新モデルが 19.6 %，更新したモデルが 11.0 %となり，モデルの更新を行うことで本人棄却率が未更新のものより約 40.0 %減少した．また他人受理率の平均はそれぞれ未更新モデルが 33.6 %，更新したモデルが 22.9 %となり，モデルの更新を行うことで他人受理率が約 30.0 %減少した．こちらはモデル更新によって性能は向上したが，元々の認識率が悪かったため話者単位のモデルよりも性能は低い結果となった．その原因としては混合数や学習データ数が少なかった可能性などが考えられる．

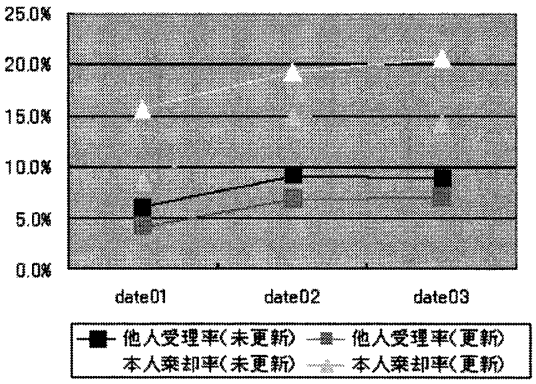

図 1. 本人棄却率と他人受理率の平均の比較

5 あとがき

本論文では携帯端末向けの話者照合によるセキュリティロックを行うため，android 端末での話者照合実験を行った．その結果，モデルを更新することで本人棄却率は未更新のモデルより約 25.0 %減少した 6.0 %となり，他人受理率は未更新モデルより約 30.0 %減少した 12.6 %となり，どちらも未更新のモデルの認識結果よりも良い結果となった．

今後，携帯端末向けに話者照合の実装を目指すためには，より照合精度を上げる必要があると考えられる．そのためには学習データを多く蓄積できる利点を生かし，より音声の変化を表現しやすいようにモデルの構造を複雑化しながら話者モデルの更新を行う方法などが挙げられる．今回，音響モデルは話者単位と音節単位で構築したが，音節単位の音響モデルは認識率があまり良い結果にはならなかった．そのため，混合数や学習データ数を増やすなど，様々な構造の音響モデルを構築することで照合精度を向上させる方法について検討していく．

参考文献

[1] 浅見太一 他，“ハフ変換による基本周波数情報を用いた雑音に頑健な話者照合”日本音響学会 2004 年春季講演論文集, 3-Q-17, pp.177-178, Mar.2004

[2] 浅見太一 他，“雑音に頑健な話者照合のための基本周波数情報の利用”情処学研報. SLP, 音声言語情報処理 pp.31-36, May. 2004

[3] 石本 祐一，“時間情報と周波数情報を用いた実環境雑音下における基本周波数推定”信学技報，vol.103, No.750, pp.49-54, Mar. 2004

[4] 松井知子 他，“話者照合におけるモデルとしきい値の更新法”信学技報, pp.21-26, Jan. 1996