# 話者認証と音声認識を統合した携帯電話のセキュリティシステムの開発

# Security system of cellular phone that integrates speaker recognition and phrase recognition

平野　邦彦
Kunihiko　Hirano
法政大学情報科学部ディジタルメディア学科
E-mail: kunihiko.hirano.zq@stu.hosei.ac.jp

Abstruct

Resentry, speaker recognition technology advances rapidly. But it has a problem that makes the other can access injustice by wiretapping. The research combines speaker recognition and Voice recognition for solving the problem. The speaker recognition and the voice recognition are act by one random word. It mounts on the cellular phone that handles a lot of individual information, and it does in solidity or more. This research is composed by the speech recognition engine of Google and the Hidden Markov Model which made from MFCC. The result, recognition rate is 85.0% . A fault other's acceptance rate is 0.01% and made others can't access injustice by wiretapping. So it is made solider.

## 1　まえがき

現在携帯電話は目覚しい発展を遂げている．それに伴い扱える情報量も増加している．ID や SUICA などの電子マネーなども扱うようになり，さらに機密資料や個人情報など外部に漏れてはならない情報も多く含まれている．

現在多くの携帯電話に搭載されているセキュリティシステムは数字 4 桁によるダイヤルロックである．しかしダイヤルロック方式には欠点がある．1 つ目にパスワードの紛失や盗難，流出があげられる．これにより使用者以外の者がパスワードを入手した場合，他人によるアクセスが可能になってしまう．2 つ目は総当り攻撃により不正アクセスが行われてしまうことだ．4 桁のダイヤルロック方式では 10000 通りの組み合わせ数が存在する．しかしこれは最高でも 10000 回アクセスを行ってしまえばセキュリティを破られ不正アクセスされてしまう．個人情報を多く所持する携帯端末には今よりも強固な堅牢性が求められている．そこで新たに注目されているのが生体認証である．

生体認証は使用者の生体を使用したものをパスワードとしているため，紛失する恐れも無い．現在生体認証には指紋，静脈，網膜認証など様々な方法がある．文献 [3] のような複数の生体認証を統合したシステムなどの先行研究も進んでいる．しかしいずれのシステムも特殊な機器を必要としそれを携帯電話で実現しようとすると，端末に負荷がかかり効率的ではない．そこで人間の声をキーとした生体認証である話者認証に注目した．話者認証ならば必要な機器は携帯電話に常備されているマイクだけであり，端末に余計な負担をかけることが無いからである．しかし話者認証にはセキュリティシステムに実装するうえで大きな欠点が存在する．それは盗聴された音声データでなりすまし認証ができてしまうことだ．

そこで本研究ではパスワードをランダムとした音声認識システムと話者認証システムを統合した認証システムを提案する．音声認識を導入することで盗聴への対策になることを次章で示す．

Supervisor: Prof. Katunobu Itou

## 2　音声を使用した認証システム

話者認証とは人間の声から個人を認識 (識別や認証) するコンピュータによる処理である．音声から特徴を抽出し，話者ごとにモデル化し，それを使って個人の声の認識を行うことである．話者認証の応用は 2 種類に分類される．1 つは，ある人物が本人の主張している通りの個人であるかを照合，認証するものである．これを話者照合と呼ぶ．もう一つは，誰だかわからない声をどの話者が発話したものかを識別するものである．これを話者識別と呼ぶ．話者照合では，話者の声を一つのテンプレートと照合すればよいが，話者識別では記憶しているあらゆるテンプレートと照合する必要がある．本研究ではユーザーと他人との区別のみが可能であればよい．これを認証するためには前者のシステムが効率よく判別できる．今回使用するのはセキュリティシステムへの実装であるため，話者照合を行うシステムを作成する．

話者照合には発話内容依存型と発話内容独立型 (以下依存型，独立型と略称) が存在する．依存型はあらかじめユーザーにテンプレートとなるパスワードを登録しておき認証時はそれを読み上げ音響特徴の照合を行うものである．独立型はユーザーにあらかじめ一定の音声を登録してもらい，その音声から話者の特徴を話者モデルとして構築，認証時はパスワードなどの定められた読み上げ文はなく，自由発話による音声の音響特徴と話者モデルとの距離で判定を行う．

発話依存型の先行研究として文献 [2] に富士通の Voice-GATE2 というシステムなどが利用されている．このシステムは音声によるパスワードを読み上げ，パスワードを話者認証するシステムである．このシステムは事前に設定しておいた暗証番号の発話を行うだけで認証ができるというものだ．しかしこのシステムにも盗聴による危険性が隠されている．事前に登録しておいた暗証番号の音声データごと盗聴されてしまえばなりすましによる不正アクセスが可能である．独立型の場合パスワードが定められていないため，音声データの盗聴の心配は無い．しかし依存型に比べ独立型は認証精度が落ちてしまう [5].

今回提案するシステムはこの依存型と独立型の中間に位置するものである．ある領域内で変動するパスワードを使用するため音声盗聴の心配も無く，精度も独立型ほど下がるわけでは無い．さらに音声認識により読み上げたパスワードが正解かを判断し正答の場合のみ認証を許可することで堅牢性の向上も図る．自分の声を認証キーとするためキーを持ち歩く必要も無く，忘却や流出してしまうことも無いという利点もある．

さらに今回実現するプラットフォームが携帯端末ということもあり，あらゆる環境下での使用状況が想定される．背景雑音による認識率の低下への影響も考慮しつつ，システム作成を行っていく．

## 3　システム概要

本研究で目指すシステムは，ランダムパスワードを使用の話者照合と音声認識の両立したセキュリティシステムだ．ランダ

ムで生成された 3 つのアルファベットをユーザーが読み上げ携帯端末に音声データとして入力する．話者照合ではユーザーを判断し，音声認識では読み上げた語句を文字として認識しアルファベットと比較する．お互いが一致することのみでユーザーにアクセス権限が与えられる強固で堅牢性の高いシステムを作成する (図 1)．

システムの概要を以下に示す．まずユーザーが認証を行う際にランダムパスワード生成が行われる．生成されたパスワードをユーザーが読み上げを行い話者照合を行い他人と登録ユーザーとの判別を行う．認証されれば次のプロセスへ，他人と判断されればプロセスをパスワードの生成まで戻す．その後，発話された音声が生成されているパスワードを正確に読み上げられているか音声認識を行い判定を行う．誤認識の場合盗聴データにアクセスと判断しプロセスを初期に戻し，認識が成功の場合のみユーザーに使用許可を与え，プロセスを終了する．以上が提案システムの概要である．

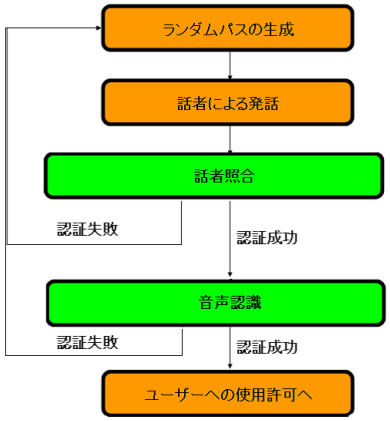


**図 1.** システム概要

### 3.1 ランダムパスワード

本研究では認証パスワードをランダムに選出されるアルファベット 3 文字のワンタイムパスワードとする．固定式のパスワードの場合，セキュリティ向上のために定期的に変更を行ったり，類推されにくい文字列を使用する必要がある．しかしワンタイムパスワードとは，パスワードを毎回異なるものにして，意味のない文字列を使用することで不正アクセスを行いにくくするものである．

現在主に使われているダイヤルロックは数字 4 ケタであり，10000 パターンである．しかし今回の試みではアルファベット 3 文字の組み合わせ，17576 パターンである．組み合わせ数が従来のシステムと同等である．アルファベット 5 文字や 7 文字とすると，さらに組み合わせパターン数が多くなり堅牢性の向上が期待できる．しかし 4 文字以上のアルファベットの羅列は一見では記憶することが難しく，また読み間違いや読み上げ中に詰まってしまい円滑に音声を読み上げることができない場合がある．

以上のことからデータは 3 文字に統一しシステム作成を行う．今回，ワンタイムパスワードを使用することで盗聴した固定音声データ文字列ではこの認証に対応することが不可能である．その結果不正アクセスに対する堅牢性を向上させる．

### 3.2 認証システム

話者照合では事前に登録されていた話者モデルと，発話によって取得された音声データとの比較を行い話者の判断を行う．目標とするシステムは認証システムであるので今回は登録されているユーザーかそれ以外の話者かを判別する．もし登録されているユーザー以外の他者が認証を行う場合この時点で発話された音声は他者と判別され，アクセス権限は取得することはできない．しかし盗聴データによる認証の場合，音声は登録されているユーザーとまったく一緒の可能性があるため不正アクセスをこの段階では防ぐことはできない．そのため，以下の音声認識を行う．

音声認識は読み上げた音声を文字データとして変換するシステムのことであり，当システムではユーザーが発話した音声が実際にランダムで提示されたパスワードを読み上げているかを判断するために使用する．ここで盗聴データでの不正アクセスの防止をおこなう．この時点でユーザーの発話が提示されたパスワードと一致しなかった場合，認証した音声データは発話ミスもしくは，盗聴された音声データである可能性がある．なぜなら実際のユーザーの場合ランダムに表示されるパスワードを読み上げることは容易である．しかし盗聴データでは取得した音声データが固定であるためランダムに生成されるパスワードに対応することは不可能であるからだ．提示されたパスワード通りに読み上げていた場合のみユーザーにアクセス権限を与える．

以上 2 つのプロセスを通し，認証を行う．

### 3.3 システム実装により向上する点

本研究では話者照合に音声認識を統合することにより盗聴データによる不正アクセス防止，それに伴う携帯端末の堅牢性の向上を目指してきた．堅牢性の向上を含めた上で当システムを実装すると以下のような利点がある．

1. 代表的な認証方法である指紋認証やダイヤルロックではタイピングや指紋読み取りなど手を用いた操作やそれを読み取る機器が必要になる．しかし，提案するシステムでは機器に触れる必要性がなく，ユーザーへの負担が軽減される．
2. 音声認識を併用することで話者照合の欠点である音声の盗聴などの不正アクセスを防ぐことができる．
3. パスワードはランダムで提示される単語であるため，ユーザー以外の音声 (盗聴による音声データ) では対応することが難しく，さらにランダムに提示される単語をキーとするため，パスワードの紛失や盗難を抑えることができる．
4. 認証時にマイクが必要となるが，携帯電話にはマイクが常備されており，新たな機器の追加は不要である．
5. 現在主流の暗証番号 4 ケタの物より組み合わせ数が多いため，端末の堅牢性を維持することができる．
6. 従来のダイヤルロック 4 ケタと比べると 4 語から 3 語に必要な語句が 1 文字削減される．これはユーザーへの負担軽減と使用効率の向上にもつながる．

## 4 システムの設計と評価

当システムでは音声認識と話者照合の二つのシステムを保有している．堅牢性の向上を考えたところ高精度の認識率を求められることになる．ここではより当システムに適応しているシステムの作成を行い，作成したシステムを他のシステムと性能評価を行うことで製作システムの優位性を証明する．

### 4.1 HMM を使用した単語単位モデルの構築

音響モデルの作成および性能評価実験をおこう．今回話者照合を行う上で 2 つの話者モデルを HTK により作成した．一つは隠れマルコフモデル (以下，HMM と略称) を使用し単語を単位とした確率モデル．もう一つは混合ガウス分布 (以下，GMM と略称) を使用し話者を単位とした確率モデルである．確率分布関数を求める際に，ガウス分布は単一ピークを持つ分布であり複雑な分布を表現することは出来ない．そのため HMM は複数の状態を持ち，各状態ごとにガウス分布を用い出力確率を構成する．そのため HMM を利用した話者モデルはモデルの単位が単語単位となる．一方混合ガウス分布 (GMM) は，各ガウス分布に重み付けを行い，一つの状態にまとめたものであるため状態数が 1 つとなり単位は話者となる．

上記の二つのモデルをそれぞれ作成し同条件下で比較，検証を行うことで高精度で当システムにより適しているシステム作成を行う．当システムではアルファベット 26 単語のうち 3 文字の発話で話者の照合を行う．実際に今回のモデル作成に用いた学習用のデータもランダムで選出されたアルファベット 3 文

字の組み合わせを 10 個用意し作成した．ユーザーの学習データは 10 文を 4 回，他人の学習データは男女 15 人で 10 文ずつを使用し，個人モデルの作成を行った．ユーザーモデルを男女 15 人で作成しその認識結果の平均を使用して精度の評価を行う．使用する実験データの取得状況は下記の表にまとめてあるものを使用した (表 1) 評価データは学習データを作成していない月の音声データを使用した．

認識率の評価としては本人棄却率と他人受理率の 2 つが存在する．本人棄却率は本人を他人と認識してしまうことであり，他人受理率は他人を本人として認識してしまうことである．当システムでは堅牢性の向上を目標にシステム作成を行っている．そのため他人受理率を優先し評価対象とする．

システムの評価結果として HMM で作成した音響モデルの本人棄却率 10% 他人受理率 0% となり GMM では混合数 16, 32, 54, 128, 256 までのモデルを作成したところどの混合数も本人棄却率 20% 他人受理率 13% となった (表 2)．この認識率を比較し検討したところ，HMM による話者モデルのほうが当システムに対して認識率が高く適していることがわかった．当システムの話者照合システムは HMM を使用した単語単位で作成された音響モデルを使用することとする．

**表 1.** 実験データ採取

| | |
|---|---|
| 録音機器 | HT-03A 実機マイク |
| 収録期間 | 3 ヶ月　月に 1 度 |
| 収録人数 | 男女 15 人 |
| 収録内容 | アルファベット 3 文字 40 文 |
| サンプリング周波数 | 16kHz |
| パラメータ | MFCC12 次元 |

**表 2.** HMM 文字単位モデルによる認識結果 (個)

| | 誤認識数 | 認識総数 | 誤り率 (%) |
|---|---|---|---|
| 本人棄却 | 16 | 150 | 11 |
| 他人受理 | 3 | 2100 | 0.01 |

### 4.2 環境音モデルの構築

上記の話者モデルに使用した音声は静寂な室内で収録されたものである．この話者モデルは同環境で収録された音声に対しては上記の性能を発揮する．しかし，収録した音声が同環境で無い場合，同じ性能を発揮することはなく認識率は著しく低下する．実際に路上（60dB）で収録した音声を上記で構築した話者モデルで認識を行うと，本人棄却率は 100% となる．そこで当実験では環境音下で話者照合を行えるよう対策を行った．

具体的に行った対策としては，上記の話者モデルの学習データに環境音下で採取した音声データを組み込み作成した．使用した音声データは駅のホーム（70dB）にて収録したものをユーザーの話者モデルに組み込み，他者モデルに対しても同条件下での学習データを統合した. さらに環境音のみのモデルも作成し，環境音と音声との差別化を図った．評価用のデータは上記で使用した評価用のデータを使用する．男性話者モデルを使用し，環境音なしの評価用データでは本人棄却，他人受理共に 0% である．

実験結果として，本人棄却率は 80% となり他人受理率は 0% と以前と変わらぬもとのなった. さらに事後的に閾値を設定し再認証を行うと本人棄却率は 20% となった．この結果から環境音への対策として当実験は有効であると考えられる．

### 4.3 音声認識システムの選定と向上

音声認識を行う上で，様々な音声認識ソフトが存在する．当システムに適したシステムを選定するために実際に使用する状況下での認証実験を行った．今回システムの性能評価を行ったのは Google の音声認識エンジン，Julius，Julian，HTK を使用した自作モデル，の 4 つである．実験状況はスマートフォンの実機マイクを使用し，それぞれのシステムに対し話者 4 名，アルファベット 3 文字，10 文の発話を行った. 結果誤り認識率が 34% と，Google の音声認識が最も精度が高くなった．(表 3)

**表 3.** 各音声認識システムの性能評価 (個)

| | Google | Julius | Julian | HTK |
|---|---|---|---|---|
| 誤り総数 | 12 | 37 | 20 | 26 |
| 平均誤り率 | 34% | 92% | 50% | 65% |

認証実験により 4 システムの中では Google の音声認識が最も精度が高いと判明した．しかし携帯の認証システムに実装するには更なる精度向上が不可欠であると考えられる．その対策として今回，話者の発話を行い，話者ごとによって発話が行いにくい単語への適応を行った．

実際に行った実験内容は，話者照合の話者モデルを構築する際取得する話者からの発話 (アルファベット) を音声認識にかける．発話された内容が決められた語句を正確に読み上げられているか否かの判別を行う．誤認識してしまった場合はユーザーにとって発話が行いにくい単語だと判断し，パスワード生成の時点でその単語を除外する．その結果，認証の際生成されるパスワードはユーザーにとって読みやすいものとなり，認識率の向上へと繋がる. 実際に話者 4 名に対し当実験を行ったところ誤認識率が 34% から 10% へと向上した．

### 4.4 多環境での音声認識精度

音声認識を行う際, 様々な環境下での実働が考えられる．雑音に対しての耐性が無い場合, 音声認識と話者照合の両方の認証によりユーザーはアクセスが可能になるため, 雑音下で認識が不可能である場合, ユーザーは認証を成功させることは無い. そのため雑音に対する耐性が求められる．当実験では Google の音声認識エンジンを使用し，それぞれの騒音を 50(室内), 60(車道)，70(駅ホーム)，80(ゲームセンター)dB と設定し雑音に対する耐性実験を行った．認識は実際にシステムで使用されるアルファベット 3 文字のパスワードで行い，各雑音下での認識率を計測した．

その結果として 50dB の環境音下では誤認識率 20%，60dB では 22%,70dB では 32%，80dB では 77% という認識結果となった（図 2)．この実験結果から，70dB 程度間での環境音では認識精度に多少の影響は与えるが，精度を大幅に落とすことは無いと考えられる．しかし環境音が 80dB を超えてしまうと認識が難しいことがわかった．これは音声認識エンジンだけの問題ではなく，自身が発話した音声がマイクに認識される間に，環境音によってかき消されてしまっている可能性がある．そのため発話者とマイクとの距離を他の状況とは変え，マイクに隣接する形で発話を行うことで認識精度を少し向上させることができるとも考えられる．しかし 80dB を超える環境音を持つものは鉄道の線路脇などのごく少数なものしかないため，当システムは日常生活下で使用する場合は問題が無いものとする.(下記の図は認識精度向上後，前).

以上の話者適応により 90% の認識率を発揮し，70dB の環境音下での使用が可能という結果から，当システムに対し Google の音声認識エンジンは有効であると考え，これを使用するもとのする．

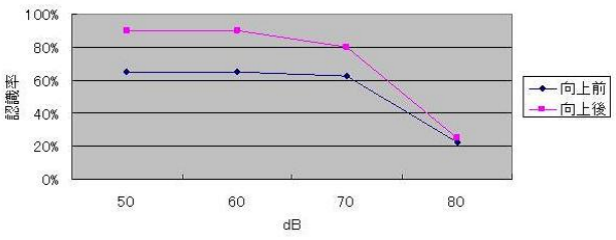


**図 2.** 環境音状況下での認識実験

## 5 システム統合による性能実験

本研究で目標とすることは話者照合と音声認識を複合することで盗聴による不正アクセスの防止し，それに伴うシステムの堅牢性の向上である．システムの実験は 2 つの方法をもって行う．当システムで使用する話者照合のモデルは前章で作成した話者照合を作成した状況と同じ実験環境，データ種類，収録人数，特徴量で行った (表 1).

1 つ目は音声認識を実装することにより盗聴データに対する耐性性能の向上評価を行う．話者照合単体のシステムと，音声認識と話者照合を統合したシステムとの性能評価を行う．以前認証に使用したアルファベット 3 文字が盗聴されたと仮定し，話者認識単体のシステムと音声認識を複合した当システムとの 2 つで，認識率を比較する．実際に行った実験としては，ランダムに生成されたアルファベット 3 文字を 50 文を盗聴データと仮定し，双方のシステムで認証を行った．話者照合単体の場合では認証を行った単語のうち，4 文以外は本人と判別をされた．しかし，音声認識を統合をした場合，盗聴によって事前に取得されたデータは固定された 3 文字であるため，認証用に使用されているアルファベットの文字列とは違うものとなる．その結果，認証されず 50 文すべてが本人の発話と判別されなかった．結果として話者照合単体の場合 92% 認識していてしまったのに対し，当システムではそれを 0% に抑えることができた。

2 つ目は音声認識の認識率と話者照合の本人棄却率の 2 つのシステムの認識率の論理積により算出し評価を行うことで，システム全体での認識率の算出を行った．その結果，全体の認証精度は話者 15 名で認証実験を行ったところ，個人差はあるものの平均として 15% の誤り認識率となった (表 4).

**表 4.** システム全体での認識性能 (誤り率 (%))

| 被験者 | 男 A | 男 B | 男 C | 男 D | 男 E |
|---|---|---|---|---|---|
| 本人棄却率 | 70 | 0 | 0 | 20 | 50 |
| 音声認識 | 10 | 20 | 10 | 0 | 0 |
| 統合認識率 | 27 | 20 | 10 | 0 | 20 |
| 被験者 | 男 F | 男 G | 男 H | 男 I | 男 J |
| 本人棄却率 | 0 | 0 | 10 | 10 | 10 |
| 音声認識 | 0 | 0 | 20 | 10 | 0 |
| 統合認識率 | 50 | 0 | 28 | 19 | 0 |
| 被験者 | 男 k | 女 A | 女 B | 女 C | 平均誤り率 |
| 本人棄却率 | 0 | 20 | 0 | 0 | 12 |
| 音声認識 | 0 | 10 | 10 | 10 | 12 |
| 統合認識率 | 0 | 28 | 10 | 10 | 15 |

## 6 あとがき

### 6.1 結論

今回の実験結果から話者照合のみで認証を行うシステムに比べ，音声認識システムを統合することによって盗聴データでの不正アクセスへの耐性が 92 ポイント向上した．この結果から，ほぼ 100% 盗聴データでの不正アクセスを防止することが可能になった. 話者照合単体での性能としては，他人受理率を 0.01% に抑えることで他人を誤って認識することは無く認証システムとしては十分な精度を発揮することができた.

様々な環境下で使用される携帯端末では，雑音への耐性が必要となる．5 章で行った実験から音声認識は 70dB 前後での環境音下で誤認識率は 20% 前後の精度を発揮した．話者照合においては，環境音下で取得した音声を学習データに使用し有効な閾値を設定することで，70dB の環境下で雑音モデル未使用の場合，本人棄却率が 100% であったのに対し，本人棄却率 20% の性能を発揮することができた．この結果から 70dB での環境音下で理論値ではあるが誤認識率 35% 前後で使用することができる．70dB は日常生活で体感することができるほぼ最大の環境音であるため，日常生活下ではほぼ全域で使用することが可能だといえる.

音声認識システムについては話者適応を行い話者に適したパスワードを生成することで，話者適応を行わないものに対して 25% ほど認識率の向上を行うことに成功した．この結果から話者適応はシステムの認識精度を向上させるためには有効な手段であるということがわかった.

音声認識と話者照合を組み合わせたシステム全体での認証精度は誤認識率 15% となった.

結果から当システムは背景雑音への耐性，盗聴データへの堅牢性の向上においては有効な精度を発揮していると言える．しかしシステム全体での誤り認証率 15% は認証システムとして低精度といえるため，今後改善が必要である.

### 6.2 今後の課題

以上のことから考えられる問題点，今後の発展を以下に記す．まず 1 つ目は全体の認証精度の悪さである．携帯電話の従来の認証システム（ダイヤルロック方式）では，ほぼ 100% 近くの精度を発揮していた. 今回の目的はシステム複合による堅牢性の向上であった. 堅牢性については従来のシステムよりも強固なものであるが, 認識率はかなりの減少してしまった. この現状の対策として考えられるものは，認識回数の増加である．1 つのパスワードに対して数回の認証の猶予を与えることにより，システム全体での認証精度を向上させようというものである．たとえば 1 回の認識に 3 回の認識数の猶予を持たすことで，理論値では誤認識率を 1% 以下にすることも可能である．認識率の向上については現在この方法を試作し向上させてく.

もう 1 つの問題点としては，雑音モデル搭載時の閾値の設定を手動で行っていることだ. 現在は他人受理が上がらず，本人棄却が最も低下する点を手動で設定している．しかし実際システムとして使用する場合は閾値の設定は手動ではできないため，自動で閾値を設定する必要がある．そのためには自動で閾値を設定できる方法を考え実際に搭載していかなければならない.

さらに今回の結果は評価用のデータが欠如していると考えられる．今回，話者照合の評価用のデータとして使用したものは話者モデルで学習データとして使用した人物と同じ人物を使用した．そのため実験としてクローズなものとなってしまった．今後としては，学習データに使用した人物以外の音声データで評価を行うことにより実験結果を信頼度を高いものにしていく必要がある.

現在，音声認識と話者照合のシステムは別々に作成し，お互いに完成しているが，統合システムでの数値を算出する作業は手作業で行っている．今後の更なる目標は，実際にモバイル端末に実現し 1 つのシステムとして実装することである.

## 参考文献


[1] 住吉貴志, 李晃伸, 河原達也,“音声認識エンジン Julius/Julian の API 実装 ” 情報処理学会研究報告. SLP, 音声言語情報処理,09196072,pp 91-96,2001-07-13

[2] 井上 美明, 熊倉 敏,” W1 音声による話者照合システム「VoiceGATEII」, および話者識別システム「VoiceSync」” IIP 情報・知能・精密機器部門講演会講演論文集,pp 1-4, 2000-03-24

[3] 小林光, 田中章浩, 岸田悟, 渡部徹, 長谷川弘,”指紋と声紋によるハイブリッド認証システムの構築” 電子情報通信学会技術研究報告. NC, ニューロコンピューティング,pp 421-426,2008-03-05

[4] 松井 知子, 古井 貞熙,”音韻・話者独立モデルによる話者照合尤度の正規化” 電子情報通信学会技術研究報告. SP, 音声,pp 61-66,1994-06-16

[5] 松井 知子,”HMM による話者認識” 電子情報通信学会技術研究報告. SP, 音声, 社団法人情報処理学会,pp 17-24,1996-01-18