学習により自己チューニング可能なRadial-Basis Function Networksによる声紋認証手法の提案

学習により自己チューニング可能な Radial-Basis Function Networksによる声紋認証手法の提案

佐藤　公信[†1]　　竹田　史章[†1]

本研究は新たな声紋認証手法の開発を目的とする．提案システムは特徴量抽出に Fast Fourier Transform を用い，識別器に未知のパターンに対して排除能力が優れた Radial-Basis Function Networks(RBFN) を用いる．提案手法は RBFN の出力細胞数を 1 としているために，特定個人の認識と未知のパターンの排除に優れると予測される．実験により認証対象者となる被験者の未学習データを評価し，提案手法の認証率および排除率を確認する．

Proposal of Voice Verification Method using Self-Tunable Radial-Basis Function Networks by Learning

Hironobu Satoh[†1] and Fumiaki Takeda[†1]

The purpose of this study is development of a new voice verification method. Fast Fourier Trasform is used for the feature extraction method in the proposed system. Radial-Basis Function Networks (RBFN), which is known that the rejection rate for the unknown pattern is hight, is used for the classifier of the proposed method. Especially, the number of the output cell of the RBFN is only one. Therefore, the proposed method excel as certification of specific person and rejection of the unknown person. In the experiment, the verification rate and rejection rate of the proposed system is confirmed using unknown pattern of subjects who are verified.

†1 高知工科大学
Kochi University of Technology

1. はじめに

現在，話者認証の研究分野では，Hidden Markov Model モデル[1] や混合ガウス分布モデルによる話者認証の研究が盛んに行われている[2),3)]．これらの手法では，認証を行うために必要とされるパラメータの作成に専門的な知識，時間およびコストが必要とされる．また，これらの手法では未知の話者を排除することが難しいと考えられる．

そこで著者らは，識別器に Radial-Basis Function Networks(RBFN)[4),5)] を用いた声紋認証手法を提案し話者認証システムの実現を目指す．識別器として用いる RBFN は，釣り鐘型関数であるガウス関数を中間層に用いているために，未知のデータに対して排除能力に優れていることで知られている[4)]．そのため話者認証システムそれ自体も未知の話者を高い確率で排除が期待される．

本手法は，専門的な知識を必要とせず，認証対象者の事前登録の実現を目指す．そこで，提案手法は音圧が閾値を越えた場合に発音部として検出するシンプルなアルゴリズムをとする．また，識別器として RBFN を用いているために，認証および排除を適切に行うためのパラメータ調整をバックプロパゲーションアルゴリズム[4)] を用いた学習により自動で行うことが可能である．さらに，特徴抽出にはスペクトル解析に一般的に使用されている Fast Fourier Transform(FFT)[6),7)] を用いる．

次に RBFN を識別器として用いた識別システムに起こり得る問題を示す．認証対象となる話者が増えることにより，音素スペクトルのバラエティが増加する．RBFN の中間層によって特徴量の近いサンプル毎にクラスター分類されるため，1 つの RBFN で相対分離を行おうとした場合には，RBFN の中間層のパラメータ数がバラエティ数と同数必要となり RBFN のパラメータは大きなものとなる．中間層細胞数が増加することによって，1 回の学習に必要な計算量の増加および認証時の RBFN の前向き計算にかかる計算量の増加といった問題が生じる．また，学習の難易度が増加し学習収束が困難になる可能性も考えられる．しかし，提案手法では 1 つの RBFN は特定の個人に特化した認証を目的としており，RBFN の出力細胞数を 1 としている．そのため，音素スペクトルのバラエティは特定個人に限定されることとなり，前述の問題回避が可能である．特定の個人と言えでも，音素によってスペクトルが異なり複数のバラエティが存在するが，RBFN の中間層の学習によってクラス分類を行うことができ，結果として提案手法は話者の認証が可能であると考える．

本論文では，声紋認証手法を提案する．また，研究の初段階として実験により認証対象および排除対象の被験者すべてを学習した RBFN を用い，被験者の未学習データを評価する

ことによって提案手法の認証率および排除率を確認する．

2. 認証手法

2.1 音声の前処理と特徴抽出

本章では提案認証手法について説明を行う．

まず，話者の発音をマイクを用いてサンプリングし wave ファイルとして保存する．wave ファイル[8] として保存された時系列のデータを無音部と話者の発音部に分ける必要がある．そこで，単位サンプル当たりの平均音圧が閾値を超えた場合，話者発音部として検出する．

次に検出された単位サンプルより RBFN に入力するための特徴量として単位サンプルに含まれる周波数スペクトルを算出する．これは次の式で表される FFT[6),7)] によって行う．

$$f_j = \sum_{k=0}^{N-1} x_k e^{-\frac{2\pi i}{n} jk} \tag{1}$$

ここで，N はサンプル数，e はネピア数，π は円周率，i は虚数を示す．FFT によって単位サンプルに含有する周波数スペクトルを算出する．FFT には一般的なハミング窓を用いる．特徴量として使用する周波数は男性および女性の発音の基本周波数範囲である 100Hz から 800Hz とする．1 回の FFT に用いるサンプル数は 4096 とする．この値を用いた場合シャノンの標本化定理[8] より 100Hz から 800Hz の音声波形の再現が可能であることが示されている．さらに，発音の音圧による変換を抑えるため，算出されたパワースペクトルをその中の最大の値で正規化したものを RBFN への入力値とする．

図 1 に作成した音声の前処理と特徴抽出を行うプログラムを示す．変換対象のサンプルが FFT によってパワースペクトルに変換されていることが確認できる．

2.2 Radial Basis Function Networks を用いた識別器

識別器には RBFN を用いる[4)]．RBFN は次の式で表される．

$$F(x) = \sum_{i=1}^{N} W_i exp\left[-\frac{1}{(2\sigma_i)^2}||x - x_i||^2\right] \tag{2}$$

ここで，N は RBFN への入力値の個数，W_i は重み，x は RBFN への入力値，x_i はガルシアン関数の中心値，σ_i は標準偏差を表している．

学習についての説明を行う．学習の前処理として RBFN のパラメータを初期化する必要がある．パラメータの初期値は次の通りである．ガルシアン関数の中心は指定する中間層細胞数の数だけ学習データとされる RBFN の入力値を初期値とする．ガルシアン関数の標準

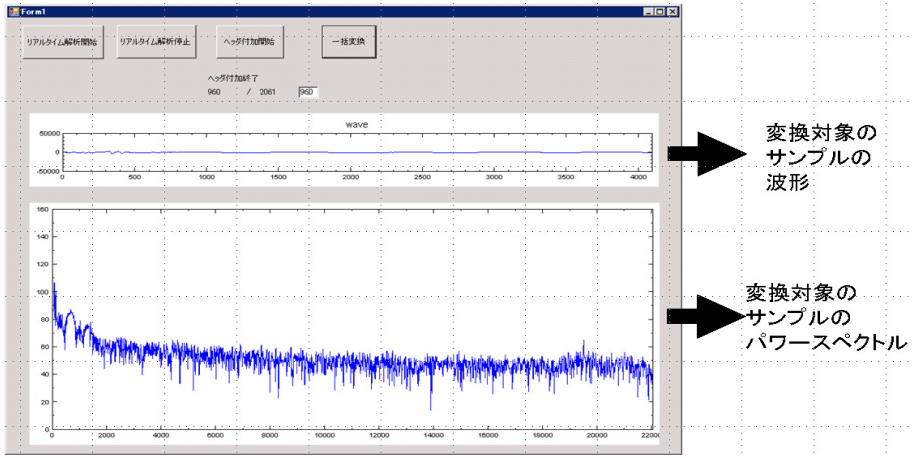

図 1 プログラムインターフェース

偏差は 0.02 として初期化する．さらに，中間層から出力層までの結合の重みは乱数によって初期化する．

学習はまず，中間層のガルシアン関数の中心位置を調整することから始める．このプロセスを行わなければ，認証対象の学習データが入力されたとしても中間層からの出力が得られず学習が進まない．ガルシアン関数の中心位置の調整は，非階層型クラスタリング手法の 1 つである k-means[4)] にて，認証対象となる話者の学習データ全てを対象として行う．ここでは，学習データをすべて使用して調整することを 1 回と定義する．終了条件は 1 回の調整において全ての中心位置の調整量が 0.3 以下になった場合とする．次に，全ての RBFN パラメータはバックプロパゲーションアルゴリズムによって調整される．ここでは，認証対象となる話者のデータおよび排除対象となる話者のデータが用いられる．認証対象の話者のデータが RBFN に入力された場合に出力値は大きく，排除対象となる話者のデータが RBFN に入力された場合に出力値は小さくなるように RBFN のパラメータが調整される．

次に RBFN の前向き計算時の出力値の扱いについて説明を行う．出力値が閾値を越えた場合には，RBFN に入力されたデータは認証対象の話者であると認証する．出力値が閾値を越えたなかった場合には，RBFN に入力されたデータは認証対象の話者ではないとして排除される．

ここまで説明を行った認証アルゴリズムのフォローを図 2 に示す．これらの処理は入力

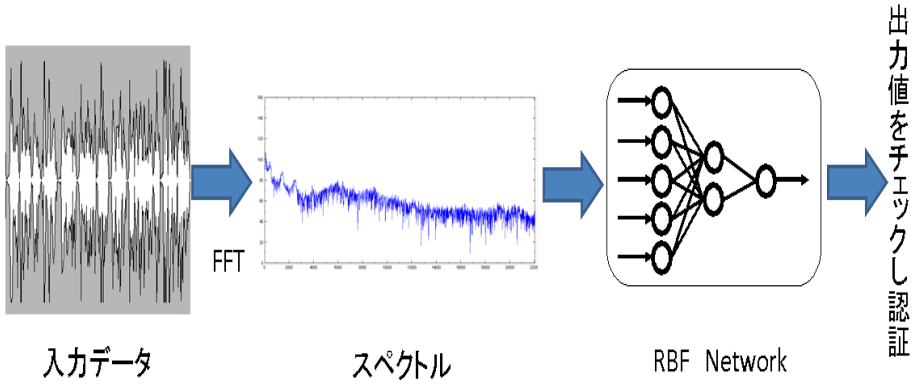

図 2 提案認証手法フロー

された wave ファイルが終了するまで一定のサンプル数検出対象となるサンプルをシフトさせ繰り返される．ここでは FFT に用いるサンプリング数を 4096，さらに移動サンプル数を 500 とした場合の発音検出対象となる wave ファイルにおけるサンプルのシフト手法のイメージを図 3 に示す．

3. 提案手法の検証実験

本章では，提案手法の検証実験を行う．被験者は被験者 A から被験者 H の男性 8 名とする．

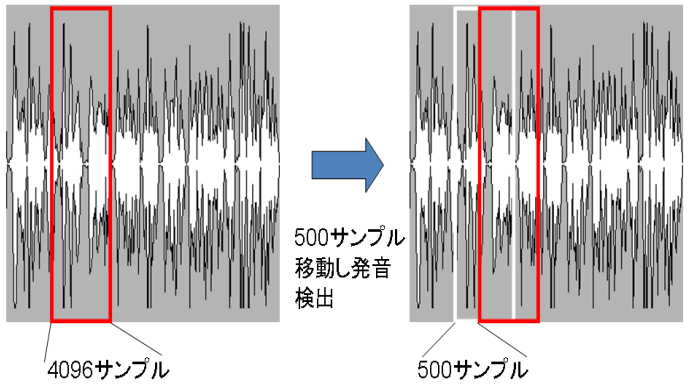

図 3 サンプリングデータのシフトに関するイメージ

図 4 被験者のマイク装着

表 1 認証対象者の認証率

	認証率 (認証数/評価数)
被験者 A	65.2% (550/844)

実験手順はまず,「フェルマーの最終定理」序章のはじめから 2 段落[9] を被験者が朗読し，マイクにて集音を行う．被験者による朗読は 2 回行う．1 回目にサンプリングした wave ファイルは学習データとして用いる．2 回目にサンプリングした wave ファイルは評価用データとして用いる．マイクを装着した被験者を図 4 に示す．マイクはヘッドセッドマイクである Logicool 社 ClearChat を使用する．集音された信号は SONY 社製 Vaio モデル PCG-6G1N に内蔵のサウンドカードに入力し，Steinberg 社 NUENDO4.3 によってサンプリングされる．サンプリングレートは 44.1KHz で，量子化ディプスは 16bit とする．また，サンプリング時には入力信号のクリッピングを防止するために waves 社 L3-LL Multimaximizer によってリミッティングを行う．さらに NUENDO のマスターフェーダ後の最終段階で Apogee 社 UV22 にてデザリングを行う．これは NUENDO のミキサーが 32bit で処理されているためである．記録されたデータは wave ファイルへと書き出される．wave ファイルは作成したプログラムを用いて 2.1 にて示した前処理および特徴抽出を行う．500 サンプルずつシ

表 2 認証対象者の認証率

	排除率 (排除数/評価数)
被験者 B	89.2% (857/961)
被験者 C	71.6% (717/1001)
被験者 D	79.0% (747/946)
被験者 E	80.3% (783/975)
被験者 F	85.0% (878/1033)
被験者 G	71.1% (700/984)
被験者 H	94.1% (1155/1228)

フトさせ発音部の検出，特徴抽出を行い RBFN へと入力するデータへと変換する．被験者 8 名の中から 1 名を認証対象者として選出し学習データとする．その後，学習データを用いて学習する．学習の主なパラメータを次に示す,

- 学習定数：0.05
- 慣性定数：0.95
- 中心値修正の比例定数：0.0005
- 標準偏差修正の比例定数　：0.00005
- 学習終了の最終誤差判定値　：0.0001
- 学習終了の最大学習回数　　：1000

中間細胞数は学習データ数から 100 を引いた値とした．

最後に認識対象者および排除対象者の評価用データを用いて認証を行う．

認証対象者の評価データを用いて算出した認証率を表 1 に示す．また，各排除対象者の評価データを用いて算出した排除率を表 2 に示す．評価数が被験者毎に異なるのは, 被験者の朗読速度の違いによって検出された発音時間が異なるためである．

実験の結果より，認証対象となる被験者の認証率は 65.2%，排除対象となる被験者の平均排除率は 81.9%となった．平均排除率が 81.9%となった原因を次に示す．図 4 に示すようにポップスクリーンを使用せずにマイクを被験者に装着したために朗読時にかぶりが発生した．これは，現実的に起こりうる現象として特に波形を編集せず実験に用いたため，かぶりの音素スペクトルが各被験者間で近い値になったと考える．また，認証率が 65.2%となった原因としては，単位サンプル毎の音圧のばらつきが大きく，FFT 後の正規化を用いてもこの差異を吸収できなかったものと推測される．

4. おわりに

本論文では新たな声紋認証手法を提案した．提案した声紋認証手法の特徴抽出は FFT を用いた．さらに識別器としては RBFN を用いた．また，RBFN の出力細胞数を 1 としたため，特定の個人の認証および未知の学習データへの排除能力も高くなると考えた．

被験者 8 名の中から 1 名を認証対象者とし提案手法の評価実験を行った．実験の結果，認証対象となる被験者の認証率は 65.2%であった．さらに，7 名の排除対象者となる被験者の平均排除率は 81.9%であった．

今後は，提案手法において個人毎の音圧の差異を吸収し認証率向上を目指す．

参 考 文 献

1) 谷萩隆嗣: 音声と画像のディジタル信号処理, コロナ社, pp.55–69 (1996).
2) 後藤真孝, 緒方淳：音楽・音声の音響信号の認識・理解研究の動向，日本ソフトウェア科学会論文誌, Vol.26, No.1, pp.4–24 (2009).
3) 松井知子, 黒岩慎吾：音声による個人認証技術の現状と展望 ，電気情報通信学会誌, Vol.87, No.4, pp.314–321 (2004).
4) Haykin, S.: Neural Networks a comprehensive foundation, Prentice hall, pp.256–317 (1998).
5) Christopher M. Bishop: Neural Networks for Pattern Recognition, Oxford Univ Press, pp.164–193 (1996).
6) 谷萩隆嗣: ディジタル信号処理と基礎理論, コロナ社, pp.40–54 (1996).
7) 青木直史: C 言語ではじめる音のプログラミング, オーム社, pp.21–54 (2008).
8) 青木直史: C 言語ではじめる音のプログラミング, オーム社, pp.1–10 (2008).
9) Simon Singh: フェルマーの最終定理, 新潮社, pp.10–11 (2006).