

## ウェーブレットに基づく 音声認証システムの精度向上にむけて

深田晋吾\* 馬場謙介† 高野茂† 井上創造‡ 安浦寛人†

\*九州大学工学部電気情報工学科

†九州大学大学院システム情報科学研究所

‡九州大学附属図書館

### 1 はじめに

本稿では離散ウェーブレット変換に基づく高速な音声認証システムの開発について述べる。局所的な時間周波数解析に有効なウェーブレット変換を用いて、多重解像度解析による音声信号の高周波成分を抽出する。次に、音声の高周波成分を用いて、音声の時間による周波数の強度変化であるスペクトログラム（声紋）を生成する。

本研究では、スペクトログラムから個人の音声特徴を抽出するために、高周波成分の局所的な共起性に基づく統計量を計算する。これを共起行列として表し、個人の音声の特徴量とする。また実験では、本人棄却率（False Rejection Rate：FRR）と他人受理率（False Acceptance Rate：FAR）による評価を行い、本システムにおける認証精度の検証を行う。

本稿は以下の構成から成る。第 2 章では、音声システムの構成を簡単に述べる。第 3 章で実験結果と考察を述べ、第 4 章で今後の予定を述べる。

## 2 音声認証システムの構成

### 2.1 ウェーブレット変換

本節ではウェーブレット変換について簡単に説明する。なお詳しくは文献 [1] を参照されたし。

信号のレベル  $j$  の低周波成分を  $c_k^{(j)}$ 、高周波成分を  $d_k^{(j)}$  と表し、それより解像度が低いレベル  $j-1$  の低周波成分を  $c_k^{(j-1)}$ 、高周波成分を  $d_k^{(j-1)}$  と表すと、次式 1,2

$$c_k^{(j)} = \sum_n p_{n-2k} c_n^{(j-1)} \quad (1)$$

$$d_k^{(j)} = \sum_n q_{n-2k} c_n^{(j-1)} \quad (2)$$

という関係が成り立つ [1]。ただし  $p_{n-2k}$ 、 $q_{n-2k}$  はウェーブレットの種類により決まる係数であり、それぞれを低周波フィルタ、高周波フィルタと呼ぶ。そしてこれらを総称してウェーブレットフィルタと呼ぶ。これにより、レベル  $j-1$  の低周波成分  $c_k^{(j-1)}$  から、1 つレベルの数の小さい（解像度の低い）レベル  $j$  の低周波成分  $c_k^{(j)}$  とレベル  $j$  の高周波成分  $d_k^{(j)}$  を求めることができる。このように、信号を高周波成分と低周波成分に分解することができる。このような解析のことを多重解像度解析と呼ぶ [1]。

### 2.2 特徴量抽出

本節では共起行列からどのように特徴量を抽出するのかを説明する。

共起行列を求めるには、以下の手順を繰り返す。音声信号の高周波成分の振幅（周波数強度）が 0 から  $N$  の範囲にあるとし、サンプリング数を  $T$  とする。  $N$  行  $N$  列の行列  $A$  をつくり、  $A$  の  $i$  行  $j$  列の要素を  $a_{i,j}$  とする ( $0 \leq i, j \leq N$ )。今、サンプリング数  $t$  番目の振幅  $i$  が、  $t+1$  番目の時に振幅  $j$  まで変化したとする ( $0 \leq t \leq T$ )。この時  $a_{i,j}$  に 1（ポイント）を加算する。これを  $t=0$  から  $t=N-1$  まで行う。

共起行列を作成すると、行列  $A$  において、頻繁に起こる振幅変化に対応する要素のポイントが高くなる。ある閾値以上である要素をホットスポットと呼び、その位置  $(i, j)$  を特徴量として用いる。

### 2.3 照合

本節では得られた特徴量からどのように照合を行うかについて説明する。まず、あらかじめ登録しておいた利用者の特徴量テンプレートと、入力された特徴量を比べる。テンプレートと入力された音声の特徴量（ホットスポットの位置）を比べ、一致していれば得点 1 を一致している数だけ与える。

Toward a fast Voice Authentication using Wavelet Transform  
\*Shingo Fukata:Department of Electrical Engineering and Computer Science, Kyushu University

†Kensuke Baba, Shigeru Takano, Hiroto Yasuura: Faculty of Information Science and Electrical Engineering, Kyushu University

‡Sozo Inoue: Library, Kyushu University

また、テンプレート作成用に使用した音声信号の特徴量とテンプレートの特徴量の照合を行い得点を算出し、その時の得点をテンプレート得点と呼ぶ。そして入力特徴量とテンプレートの特徴量を比較した時に得られた得点をこのテンプレート得点で割った値を得点率する。

さらに入力特徴量の総数（特徴量数）のうちテンプレートの特徴量といくつが一致したのか（一致率）を計算し、この一致率と得点率がある閾値以上であれば受理し、以下であれば棄却する。

### 3 実験結果と考察

本章では実験結果を示し考察を行う。なお音声解析にはドベシのウェーブレット [1] を使用した。また実験に使用した音声信号のサンプリングレートは 16kHz、量子化ビット数は 16bit である。音声信号にウェーブレット変換を 10 回を行い、レベル 1~10 の各高周波成分から共起行列を作成し、特徴量を抽出し照合を行った。

#### 3.1 実験結果

表 1 に実験結果を示す。10 人の被験者 A~J に「こんにちは」と 10 回発声してもらい録音し、5 個をテンプレート作成に用い、残りの 5 個を照合に使用した。テンプレート得点はテンプレート作成用に使用した 5 個の音声信号とテンプレートを照合した時の得点の平均値を使用する。得点率の閾値を 50%、一致率の閾値を 60% として照合を行った。

表 1 の名前は 10 人の被験者 A~J を表し、特徴量数は各被験者の照合用の 5 個の音声の平均特徴量数を表す。得点率と一致率は、各被験者のテンプレートと各被験者自身の音声データ 5 個との照合を行った際の得点率および一致率の平均値を表す。FRR は各被験者本人のテンプレートと 5 個の音声データとの照合を行った結果を表す。同様に FAR は各被験者のテンプレートとその他の被験者の音声データとの照合を行った結果を表す。また FAR を算出する際には、各被験者のテンプレート作成用の音声信号 5 個も照合に使用した。

表 1 より、FRR の平均は 16%、FAR は 17% という結果が得られた。

表 1: 実験結果

名前	特徴量数	得点率	一致率	FRR	FAR
被験者 A	490	89%	74%	0%	21%
被験者 B	130	66%	69%	0%	21%
被験者 C	443	95%	80%	0%	15%
被験者 D	463	91%	63%	60%	16%
被験者 E	149	64%	83%	20%	16%
被験者 F	67	51%	63%	20%	19%
被験者 G	92	83%	66%	20%	9%
被験者 H	87	71%	64%	20%	23%
被験者 I	130	72%	67%	0%	16%
被験者 J	81	62%	72%	0%	18%

#### 3.2 考察

この実験より、各被験者で FRR、FAR に大きな差が出ていた。各被験者で特徴量数が大きく異なっていたことが原因だと考えられる。

このことから共起行列を求める際に正規化をしてから特徴量を求める必要があると考えられる。共起行列の要素の最大値が 1 となるようにすることで正規化できると考えられる。

### 4 今後の予定

特徴量抽出のアルゴリズムを改良し、既存ウェーブレットフィルタの種類を変え精度を調査する。

その後音声認証に最適なウェーブレットフィルタの設計法の研究を行う。そして設計したウェーブレットフィルタによって既存のフィルタとどの程度精度に違いがあるのかを検証する。

### 参考文献

- [1] 中野宏毅, 山本鎮男, 吉田靖夫, ウェーブレットによる信号処理と画像処理, (社) 共立出版株式会社, 東京, 1999.