

# リフティングウェーブレットフィルタに基づく 音声認証システムの開発

深田晋吾\* 高野茂† 岡田義広† 藤崎清孝†

\*九州大学大学院システム情報学府情報知能工学

†九州大学大学院システム情報科学研究院

## 1 はじめに

本稿では、リフティングウェーブレットを用いた音声認証に適したフィルタの設計方法について述べる。本手法では、音声の特徴抽出として局所的な時間周波数解析に有効であるウェーブレット変換を用いる。具体的には、ウェーブレット変換を用いて入力音声信号の多重解像度解析を行い、得られた低周波成分から LPC ケプストラムを計算し、これを音声特徴量とする。このとき、自由パラメータをもつリフティングウェーブレットフィルタを再帰的に適用することによって、個人認証に最適となるパラメータを決定し、音声認証に有効な特徴を抽出するウェーブレットフィルタを遺伝的アルゴリズムを用いて実験的に求める。

## 2 音声認証システムの構成

### 2.1 リフティングウェーブレット

#### 2.1.1 ウェーブレットフィルタ

以下の式

$$c_i^0 = \sum_l h_l^o c_{2i+l}^1 \quad (1)$$

$$d_i^{(0)} = \sum_l g_l^o c_{2i+l}^1 \quad (2)$$

という関係が成り立つとする [1]。式 (1), (2) をそれぞれ低周波分解公式および高周波分解公式とよび、 $c_i^0$  は原信号の低周波成分、 $d_i^{(0)}$  は原信号の高周波成分を表す。また、 $h_l^o, g_l^o$  をそれぞれを低周波、高周波フィルタとよび、これらを総称してウェーブレットフィルタとよぶ。

#### 2.1.2 リフティングウェーブレットフィルタ

次の4つのフィルタ組

$$\begin{aligned} h_l &= h_l^o, & g_l &= g_l^o - \sum_m s_m h_{l-2m}^o, \\ \tilde{g}_l &= \tilde{g}_l^o, & \tilde{h}_l &= \tilde{h}_l^o + \sum_m s_{-m} \tilde{g}_{l-2m}^o, \end{aligned} \quad (3)$$

をリフティングウェーブレットフィルタ [1] とよぶ。ただし、パラメータ  $s_m (m = 0, \pm 1, \pm 2, \dots, \pm M)$  は任意の実数値である。

原信号に、式 (3) を適用して、以下のような新しい低周波成分  $\hat{c}_i^0$  と高周波成分  $\hat{d}_i^0$  が得られる

$$\begin{aligned} \hat{c}_i^0 &= \sum_l h_l^o c_{2i+l}^1, \\ \hat{d}_i^0 &= d_i^0 - \sum_m s_m c_{i+m}^0 \end{aligned}$$

本研究の狙いは、自由パラメータ  $s$  を調整することにより、個人の音声特徴に適應するウェーブレットフィルタを設計することである。

### 2.2 LPC ケプストラム (LPCC)

ケプストラム (cepstrum) は、信号の短時間振幅スペクトルの対数の逆フーリエ変換として定義される。

ケプストラムの特殊なものとして、線形予測分析 (LPC) によって推定された全極型のスペクトルモデル

$$H(z) = \frac{1}{1 + \sum_{i=1}^M \alpha_i z^{-i}} \quad (4)$$

が得られているとする。ただし、 $M$  はモデル数である。この時、式 (4) のシステム関数に対応するケプストラムを得る為に、フーリエ変換、対数変換、逆フーリエ変換をそれぞれ、両側  $z$  変換、複素対数変換、両側  $z$  逆変換で置き換えることによって、得られる複素ケプストラムを  $c(m)$  で表すと、次の再帰式が得られる。

$$c(m) = \begin{cases} -\alpha_1, & m = 1 \\ -\alpha_m - \sum_{k=1}^{m-1} \binom{k}{m} c(k) \alpha_{m-k}, & 1 < m \leq M \\ -\sum_{k=1}^{m-1} \binom{k}{m} c(k) \alpha_{m-k}, & M < m \end{cases} \quad (5)$$

SPEAKER IDENTIFICATION SYSTEM USING LIFTING WAVELETS FILTER

\*Shingo Fukata: Graduate School of Information Science and Electrical Engineering, Kyushu University

†Shigeru Takano, Yoshihiro Okada, Kiyotaka Fujisaki: Faculty of Information Science and Electrical Engineering, Kyushu University

式(5)で表現されるケプストラムを、線形予測モデルによるケプストラムという意味で、LPC ケプストラムとよぶ。

### 2.3 特徴量抽出 (MBLPCC)

ウェーブレット分解公式により、元の音声信号を高周波成分と低周波成分の2つに分解することができる。得られた低周波成分から LPCC 係数を求め、特徴量とする。さらにその低周波成分をウェーブレット変換により高周波成分と低周波成分に分解し、上記と同様の操作を繰り返す。こうして得られた LPCC と元の LPCC を組み合わせた特徴量を Multi Bands LPCC(MBLPCC)[2] とよぶ。

### 2.4 混合ガウスモデル

混合ガウス分布  $p(\mathbf{x}|\lambda)$  は、ガウス分布の線形重ね合わせとして以下のように表わされる[2]。

$$p(\mathbf{x}|\lambda) = \sum_{i=1}^M p_i b_i(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

ただし、 $\mathbf{x}$  は  $N$  次元のランダムベクトル、 $M$  は混合要素数、 $b_i(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$  はガウス分布、そして  $p_i$  は混合係数であり、 $\sum_{i=1}^M p_i = 1$  を満たす。ガウス分布  $b_i(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$  は以下の式で表すことができる。

$$b_i(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \frac{1}{(2\pi)^{N/2} |\boldsymbol{\Sigma}_i|^{1/2}} \times \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right\}$$

ただし、 $\boldsymbol{\mu}_i$  は平均ベクトル、 $\boldsymbol{\Sigma}_i$  は共分散行列である。

混合ガウス分布は、平均ベクトル、共分散行列、混合係数のパラメータで表される。これらのパラメータを以下のように表記する。

$$\lambda = \{p_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}$$

音声認証を行うために、 $S$  人の話者集合を  $\{\lambda_1, \dots, \lambda_S\}$  とする。判定を行う場合、テストパターンを  $(\mathbf{x}_1, \dots, \mathbf{x}_T)$  とすると、以下のように尤度が最大となる話者モデルを求める。

$$\hat{S} = \arg \max_{1 \leq k \leq S} \sum_{t=1}^T \log p(\mathbf{x}_t | \lambda_k)$$

## 3 実験

本実験では、PASL-DSR[3] 日本語音声コーパスを利用した。この中に含まれる 10 人の話者(女性 6 人、男性 4 人)のサンプリング周波数 16kHz、量子化数 16bit でサンプリングされた音声を使用した。

Haar のウェーブレットフィルタを初期段階の  $h_l^o, g_l^o$  として使用し、認証精度を最大にするリフティングパラメータ  $s_m (M = 4)$  を遺伝的アルゴリズムを用いて求めた。

本実験においては、LPCC、MBLPCC[2] と提案手法の認証精度比較実験を行った。表 1 にテスト音声時間の違いによる精度比較を、表 2 にテスト音声に雑音を付加した場合の精度比較結果を示す。

時間 [s]	LPCC[%]	MBLPCC[%]	提案手法 [%]
1.0	96.8	98.6	100
0.5	89.2	93.3	99.7
0.2	73.2	81.6	94.4

表 1: テスト音声時間による精度比較

SNR[dB]	LPCC[%]	MBLPCC[%]	提案手法 [%]
30	53.1	91.3	98.9
20	22.5	31.4	72.5
15	13.3	17.6	44.2

表 2: 雑音状況における精度比較 (テスト時間:1.0[sec])

## 4 今後の予定

被験者数が不十分であると考えられるので、被験者数を増やして実験を行いたい。

## 参考文献

- [1] S. Mallat, "A Wavelet Tour of Signal Processing (2nd edition)," *Academic Press*, 1999.
- [2] C.Hsieh, E.Lai, and Y.Wang, "Robust Speaker Identification System Based on Wavelet Transform and Gaussian Mixture Model," *Journal of Information Science and Engineering* 19, 2003, pp.267-282.
- [3] S.itabashi, "Spoken Language and the DSR Project Corpus(PASL-DSR)," 1991.(in Japanese)