

リフティングウェーブレットによる 音声認証に適したフィルタの設計

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

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

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

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^0 、 g_l^0 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

$$\begin{aligned} h_l &= h_l^0, \\ g_l &= g_l^0 - \sum s_m h_{l-2m}^0, \\ \tilde{h}_l &= \tilde{h}_l^0 + \sum_m s_{-m} \tilde{g}_{l-2m}^0, \\ \tilde{g}_l &= \tilde{g}_l^0 \end{aligned} \quad (3)$$

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 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

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

逆に、新しい低周波成分 \hat{c}_i^0 と高周波成分 \hat{d}_i^0 から、リフティング復元フィルタを用いて原信号 \hat{c}_i^1 を復元することができる。

$$\hat{c}_i^1 = \sum_l \tilde{h}_l \hat{c}_{(i-l)/2}^0 + \sum_l \tilde{g}_l \hat{d}_{(i-l)/2}^0$$

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

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

Speaker Recognition System Based on Lifting Wavelet Transform

*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

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。

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

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 [2] とよぶ。

2.3 特徴量抽出

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 LPC ケプストラム係数を求め、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。こうして得られた特徴量を Multi channel LPC ケプストラム (MCLPCC)[3] 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[3]。

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

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、 $b_i(\mathbf{x}|\boldsymbol{\mu}_i, \Sigma_i)$ 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 $b_i(\mathbf{x}|\boldsymbol{\mu}_i, \Sigma_i)$ は以下の式で表すことができる。

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

ただし、 $\boldsymbol{\mu}_i$ は平均ベクトル、 Σ_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。

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

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

入力音声信号に対して、様々な自由パラメータによるリフティングウェーブレット変換を適用し、それぞれの認証精度を検証する。

実験の詳細及び結果は発表当日に示す。

4 今後の予定

リフティングウェーブレットにより得られたパラメータを個人を認証する際の「鍵」として利用し、IC カード等を組み合わせた新しい複合認証サービスへ応用することも考えている。

謝辞

本研究の一部は、平成 21 年度戦略的情報通信研究開発推進制度 (SCOPE) 若手 ICT 研究者育成型研究開発「専門家の知識と群衆の叡智を融合する次世代図鑑検索システムの研究開発」によった。

参考文献

- [1] S. Mallat, "A Wavelet Tour of Signal Processing (2nd edition)," *Academic Press*, 1999.
- [2] K. Takao, "Cepstral and Mel-Cepstral Analysis of Speech," *IEICE technical report. Speech*, 1998, pp.33-44.
- [3] 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.