ISS-P-017

原音声波形と残差波形からの MFCC と相対位相による話者認識の比較

山本　滉己†　　山本　一公†　　中川　聖一†
† 中部大学情報工学科

1. はじめに

テキスト独立型話者認識法として、混合ガウス分布モデル(GMM)が従来から用いられてきた。また、代表的特徴パラメータであるメルケプストラム係数(MFCC)の後処理として、i-vector[1]を抽出して用いるのが最近の中心技術になっている。我々は、MFCC のような振幅スペクトラム情報と対をなす位相スペクトラムから得られる相対位相情報を MFCC と併用することにより話者認識に大きな改善が得られることを示してきた[2][3]。これは、相対位相情報が音源波形の特徴を捉えているためと考えられる。そこで、本稿では、より直接的に原音声波形の線形予測分析で得られる残差波形から相対位相情報をもとめ、話者認識に有効であるかを検討したので報告する。

2. 相対位相情報

例えば、16kHz でサンプリングされた音声波形から 256 個のサンプル点を切り出し、窓かけ後フーリエ変換すると 128 個の振幅スペクトルと位相スペクトルからなる次式の線スペクトルを得る。

$$\sqrt{X^2(\omega,t)+Y^2(\omega,t)}\times e^{j\theta(\omega,t)} \qquad (1)$$

上式の位相を表す $\theta(\omega,t)$ は、同じ角周波数 ω でも切り出す位置によって位相が変わるのを防ぐため、ある周波数の値を基準値に固定し、他の周波数における位相を基準値からの相対値として求める[2]。また、ピッチによる位相の変動を防ぐために、疑似ピッチ同期で窓を切りだす。さらに、位相の値 θ を $\{\cos\theta, \sin\theta\}$ に変換し、座標値を特徴パラメータとして用いる。

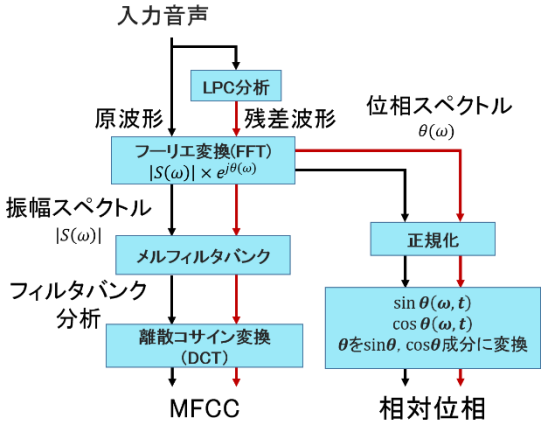

図1. 本実験で用いた特徴パラメータ抽出のブロック図

3. 話者認識実験

話者認識実験は、文献[4]と同じく、新聞読み上げ音声データベース JNAS を用いた。男女各 135 名（計 270 名）からなる各話者の 100 話文のうち、10 発話を各話者の GMM モデルの学習に用い、他の 90 発話をテスト用とし、1 文（約 4 秒の発話長）ごとに話者認識を行った。GMM の混合数は 128 とした。

音声分析条件は、16kHz サンプリング、分析窓長は MFCC は 25ms、相対位相は 12.5ms、フレームシフト長は両者 5ms とした。MFCC 抽出のためのフィルタバンクのフィルタ数は 22 個、MFCC は ΔMFCC、ΔΔMFCC、Δパワー、ΔΔパワーを含めた 38 次元である。一方、相対位相特徴は、低次の 19 個または 30 個の位相スペクトルから抽出される 38 次元または 60 次元である。

4. 話者認識実験結果

話者認識実験結果を表1に示す。原波形より抽出した MFCC、相対位相の認識率は残差波形より抽出した MFCC、相対位相の認識率よりも認識性能が良かった。

表 1　話者認識実験結果（%）

特徴パラメータ	原波形	残差波形
MFCC	99.17	97.73
相対位相(38次元)	93.22	92.64
相対位相(60次元)	95.95	94.47

MFCC と 60 次元数の相対位相の組み合わせた場合の話者認識実験結果を表2に示す。原波形より抽出した MFCC に原波形と残差波形の相対位相を組み合わせることで、ともに認識率の向上が見られた。

表 2　MFCC と 60 次元の相対位相の組み合わせた話者認識実験結果（%）

特徴パラメータの組み合わせ	認識率
MFCC(原波形)　+ 相対位相(原波形)	99.51
MFCC(原波形)　+ 相対位相(残差波形)	99.48
MFCC(残差波形) + 相対位相(原波形)	98.90
MFCC(残差波形) + 相対位相(残差波形)	98.85

謝辞　本研究は JSPS 科研費 16K12641 の助成を受けた。

参考文献
[1] 小川、塩田：i-vector を用いた話者認識、日本音響学会誌、Vol.70, No.6, pp.332-339, 2014
[2] 大塚、王、中川：話者認識における位相情報の改善、日本音響学会講演論文集、pp.213-214, 2007.9
[3] 嶋田、山本、中川：話者認識のための位相特徴抽出法の改善、日本音響学会講演論文集、pp.285-286, 2010.3
[4] 王、中川：MFCC と位相情報を用いた雑音環境下での JNAS データベースにおける話者認識の評価、日本音響学会講演論文集、pp.287-290, 2010.3