周波数領域指定マッチングの Prony ESD への適用

4E－1

松永 光輝*　中央大学大学院理工学研究科
鈴木 寿†　中央大学理工学部

1　はじめに

本研究では，過渡部の特徴量を正確に分離抽出するため，FFT に代り Prony の提案した Prony ESD (Energy Spectral Density)[1]を使用する．現在，発声された音声からその特徴を抽出するための手法として，FFT (Fast Fourier Transform)，またはその応用を用いてスペクトルを算出し特徴を抽出するという手法が主に用いられている．ここで，FFT は日本語の子音にある過渡部などの高周波域に発生する特徴量を正確に抽出することは困難であるという問題がある．Prony ESD は FFT より正確に過渡部の抽出ができるという事が分かっている[2]．しかし，Prony ESD は高周波域の特徴量に対して敏感なので，同じ音声を発声しても ESD が安定しない．その為，一般的なユークリッド距離を使用した従来のマッチング法を使用することができない．

以下は ESD ではなく周波数領域に着目した距離関数を提案し，その有効性について検証する．

2　Prony ESD について

Prony ESD は，まず採取された時系列データを，ある振幅・位相・減衰率・周波数からなる指数関数により曲線近似する周波数変換である．その算出には，まず採取された時系列データ x_i の近似値 $\hat{x}_i$ を以下のように定義する．近似値 $\hat{x}(t)$ は

$$\hat{x}(t) = \sum_{m=1}^{p} A_m \exp(\alpha_m |t|) \exp\left(j[2\pi f_m t + \theta_m]\right) \quad (1)$$

である．ここで，A_m は振幅，θ_m は位相，α_m は減衰率，f_m は周波数とする．

式 (1) は有限より，$\hat{x}(t)$ の周波数変換 $\hat{X}(t)$ は，フーリエ変換を元とし

$$\hat{X}(f) = \sum_{m=1}^{p} A_m \exp(j\theta_m) \frac{2\alpha_m}{\exp\left[\alpha_m^2 + (2\pi[f - f_m])^2\right]} \quad (2)$$

となる．よって求める ESD は

$$\hat{\zeta}_{PRONY}(f) = |\hat{X}(f)|^2. \quad (3)$$

* 中央大学理工学部，〒112-8551 東京都文京区春日 1-13-27. Department of Science and Engneering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan. E-mail: matunaga@suzuki-lab.ise.chuo-u.ac.jp
† 中央大学理工学部 (中央大学理工学研究所)

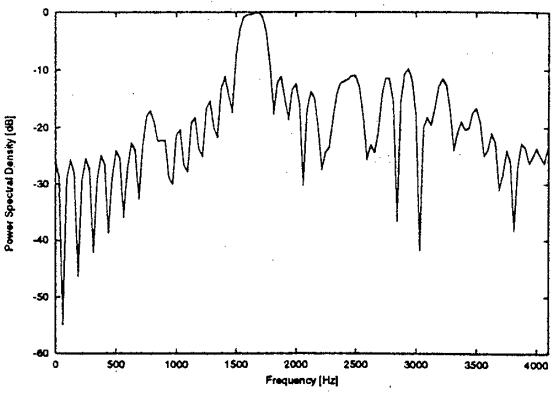

図 1. FFT による Power Spectral Density

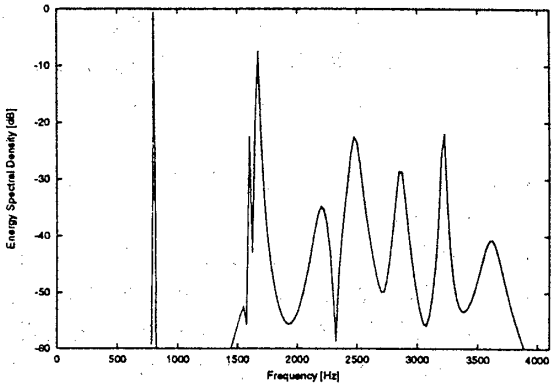

図 2. Prony Energy Spectral Density

図 1, 2 に Prony ESD の適用例を示す．例に使用した時系列データは，3 つの正弦波と white Gaussian による filtering によって得られたノイズを含んでいるものを使用した．図 1,2 により，Prony ESD では FFT よりもスペクトルを正確に描けることがわかる．

3　周波数域に着目した距離関数の提案

上述の通り，Prony ESD は，高周波域の特徴量に対して敏感である．それゆえ，同じ話者が発した声であっても，ESD の値は大きく変動する．しかし，ピーク点の現れる周波数域はほぼ一定の値をとるので，それを利用した周波数域に着目した距離関数を提案する．以下にその算出方法を示す．

まず，ピーク点における周波数の最小距離 $r(i)$ は次式のように与えられる．

$$r(i) = \min\{|P_b(j) - P_a(i)|\}. \quad (i = 1, \cdots, I) \qquad (4)$$

ただし，$P_a(i)$ を入力波形のピーク点での周波数，$P_b(j)$ を標準パターン波形のピーク点での周波数とし，i, j はそれぞれのピーク点の個数とする．式 (4) で，i を固定し，j を動かす事によって得られる $r(i)$ は入力波形に対しての標準パターン波形とのピーク点の対称，又は非対称な対応をもつ値となる．
$r(i)$ により，距離 FT は，

$$FT = \frac{1}{I}\sum_{n=1}^{I} r(i) \qquad (5)$$

となる．

4 シミュレーション

本研究のシミュレーションは，不特定な話者の発声した子音 /pa/, /pi/, /pu/, /pe/, /po/ に対して，Prony ESD を適用する．マッチング法としては，擬似的な端点フリー DP マッチングを使用し，今回提案した距離関数とユークリッド距離関数との比較をおこなった．
今回のシミュレーションにおいてのサンプルデータは，以下の条件の元で採取した．

- 単一方向性及び低域通過フィルタを搭載マイクロフォンを使用した．
- 採取条件として 16bit,8KHz, モノラルとする．
- データは，12 人から各子音を 10 回ずつ，完全に区切って発音してもらい，合計 600 個を使用した．
- フレーム周期は 8ms，窓関数にはハミング窓を用いた．
- 標準パターン波形は今回のデータ以外の話者の発声したものを使用した．

表 1 はその実験結果，図 3 はユークリット距離関数と提案手法での 5 音の認識率のグラフで，破線がユークリッド距離であり，実線が提案手法での認識率である．

図 3 より，一般のユークリッド距離を使用したものでは，中舌母音を含む /pu/，そして後舌母音 /o/ をもつ /po/ が，それぞれ 86%であるが，前舌母音 /i/, /e/ を含んでいるものは，50%，58%と低い認識率を示した．5 つの子音の中で，最も結果が思わしくないものが，後舌母音を含む /pa/ で，その認識率は 3%と極端に落ち込んでいる．この理由としては，/pa/, /po/ の標準パターン波形が類似しており，それが干渉して /pa/ の音を /po/ として認識しているためだと考えられる．
全体的な認識率を見ても 56%弱と低く，これにより，Prony ESD を適用したデータに対しては，主にスペクトルの

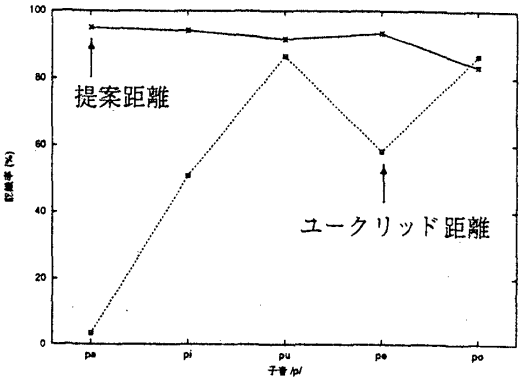

図 3. 2 手法での認識率の推移

形を重視する一般的なユークリッド距離での認識は困難であると言える．

逆に今回提案した距離関数を用いたものでは，その認識率は，/po/ が 83%とやや低いが，それ以外の /pa/, /pi/, /pu/, /pe/ では，ユークリッド距離と同じ標準パターン波形を使用しているにも関わらず，それぞれ 95%, 94%, 91%, 93%と高い認識率を示した．これは，この提案手法が，ESD よりも周波数を重視しているため，上述の /pa/,/po/ のような類似しているスペクトルをもつ波形に対してその影響を受けることが少ないことを示している．全体の平均認識率でも，ユークリッド距離での認識率より，約 30%弱向上している．

この結果，Prony ESD のようなスペクトルのピークが鋭く現われる特徴抽出法には，今回提案したような周波数帯域を重視した距離関数が有効だと分かる．

5 おわりに

本論文では，Prony ESD の特性に対して，新たな距離関数を提案し，従来の距離関数とのシミュレーションによる比較検証をおこなった．
シミュレーションの結果，今回提案した新たな距離関数は，Prony ESD に対しては従来の距離関数よりも有効であると言えた．

今後の課題として，既存の特徴量抽出手法 (ケプストラム，LPC 分析，LPC ケプストラム) との認識率による比較シミュレーションをおこなう必要がある．

参考文献

[1] S.M. Kay, S.L. Marple, "Spectrum analysis - a modern perspective," *Proceedings of the IEEE*, vol.69, November 1981

[2] 松永, 鈴木, "Prony ESD の子音認識に関する一考察," 第 *69* 回情報理論とその応用シンポジウム, pp647-650, December 1998.