2 次元 AR モデルと高速 RLS 法による音声分析

3R－4

堀田 英輔† 西川 清† 宮永 喜一††

† 金沢大学 工学部 電気・情報工学科
†† 北海道大学 工学部 電子情報工学専攻

1 まえがき

音声信号を確率モデルを用いて分析し，音声合成や符合化，認識などへ応用することが行われている．また，これらの手法を雑音が存在する環境で用いるため，複数のマイクロホンを用いて雑音除去を行う研究もなされているが，音声の到来方向が既知であると仮定する必要があった [1]．本稿では直線上に等間隔に配置されたアレーマイクロホンの位置を空間座標軸にとり，音声信号が時間と空間の 2 次元信号であると考える．このとき，アレーマイクロホンに異なる角度で入射する 2 種類の信号の 2 次元周波数平面上でのスペクトルは理想的には分離できるため，音声の到来方向の情報を予め必要とせずに，入射角度の異なる音声と雑音のスペクトルを分離して推定することができる．本稿では 2 次元非対称半平面 AR モデルを用いてそれらのスペクトルを適応的に推定する手法について提案する．

2 RLS に基づく係数更新式

逐次最小 2 乗法 (Recursive Least-Squares: RLS) の評価関数を次式で定義しよう．ただし，アレーマイクロホン信号 yの 2 次元離散領域での時間座標を iT (Tはサンプリング周期)，空間座標を jD (Dはマイクロホン間隔) とする．すなわち

$$y(iT, jD) = y(i,j)$$

と表記する．また，2 次元 AR モデルの係数存在領域として最も一般的な非対称半平面を考え [2]，さらに信号の時間的因果性を仮定する．

$$J(k) = \sum_{i=1}^{k} \sum_{j=l}^{L-l+1} \lambda^{k-i}\{y(i,j) - \hat{y}(i,j)\}^2 \qquad (1)$$

$$\begin{aligned} \hat{y}(i,j) &= -\sum_{q=1}^{l-1} a_{0,q}(k) y(i,j-q) \qquad (2) \\ &\quad -\sum_{p=1}^{n-1} \sum_{q=-l+1}^{l-1} a_{p,q}(k) y(i-p, j-q) \\ &= \boldsymbol{p}^T(k)\boldsymbol{h}(i,j) \end{aligned}$$

$$\begin{aligned} \boldsymbol{p}^T(k) = [\, & a_{0,1}(k)\, a_{0,2}(k) \cdots a_{0,l-1}(k) \\ & a_{1,-l+1}(k)\, a_{1,-l+2}(k) \cdots a_{1,l-1}(k) \cdots \\ & a_{n-1,-l+1}(k)\, a_{n-1,-l+2}(k) \cdots a_{n-1,l-1}(k) \,] \end{aligned}$$

$$\begin{aligned} \boldsymbol{h}^T(i,j) = [\, & -y(i,j-1) - y(i,j-2) \cdots \\ & -y(i,j-l+1) - y(i-1,j+l-1) \\ & -y(i-1,j+l-2) \cdots - y(i-1,j-l+1) \cdots \\ & -y(i-n+1,j+l-1) - y(i-n+1,j+l-2) \\ & \cdots - y(i-n+1,j-l+1) \,] \end{aligned}$$

式 (1) の評価関数は，時間に関しては過去の信号の影響を忘却させる指数重み係数λ ($0 < \lambda \leq 1$) が導入され，prewindowing が仮定されている．そして，空間に関しては $y(i,1)$ から $y(i,L)$ までを用いる共分散法が仮定されている．このとき，式 (1) を AR 係数 $a_{p,q}(k)$ に関して最小化して得られた正規方程式を逆行列の補助定理を用いて適応的に解くと次のアルゴリズムが得られる．

[2 次元 AR モデルの係数推定の基本更新式]

$$\hat{\boldsymbol{p}}^{(l-1)}(k) = \hat{\boldsymbol{p}}(k-1),\ \boldsymbol{F}_n^{(l-1)}(k) = \lambda^{-1}\boldsymbol{F}_n(k-1)$$

$$J_{min}^{(l-1)}(k) = \lambda J_{min}(k-1)$$

$for\ m = l, l+1, \cdots, L-l+1$

$$\nu(k,m) = y(k,m) - \hat{\boldsymbol{p}}^{(m-1)T}(k)\boldsymbol{h}(k,m)$$

$$\boldsymbol{G}_n^{(m)}(k) = \frac{\boldsymbol{F}_n^{(m-1)}(k)\boldsymbol{h}(k,m)}{1 + \boldsymbol{h}^T(k,m)\boldsymbol{F}_n^{(m-1)}(k)\boldsymbol{h}(k,m)}$$

$$\hat{\boldsymbol{p}}^{(m)}(k) = \hat{\boldsymbol{p}}^{(m-1)}(k) + \boldsymbol{G}_n^{(m)}(k)\nu(k,m)$$

$$\boldsymbol{F}_n^{(m)}(k) = \boldsymbol{F}_n^{(m-1)}(k) - \boldsymbol{G}_n^{(m)}(k)\boldsymbol{h}^T(k,m)\boldsymbol{F}_n^{(m-1)}(k)$$

$$f(k,m) = y(k,m) - \hat{\boldsymbol{p}}^{(m)T}(k)\boldsymbol{h}(k,m)$$

$$J_{min}^{(m)}(k) = J_{min}^{(m-1)}(k) + \nu(k,m)f(k,m)$$

$$\hat{\boldsymbol{p}}(k) = \hat{\boldsymbol{p}}^{(L-l+1)}(k),\ \boldsymbol{F}_n(k) = \boldsymbol{F}_n^{(L-l+1)}(k)$$

$$J_{min}(k) = J_{min}^{(L-l+1)}(k)$$

ここで，$J_{min}(k)$ は式 (1) の最小値，すなわち

$$J_{min}(k) = \sum_{i=1}^{k} \sum_{j=l}^{L-l+1} \lambda^{k-i}\{y(i,j) - \hat{\boldsymbol{p}}^T(k)\boldsymbol{h}(i,j)\}^2 \quad (3)$$

である．

3 FRLS に基づく係数更新式

本稿で考えている AR モデルは 2 次元であり，その係数総数は比較的多いと考えられるため，第 2 節のアルゴリズムを高速化する必要がある．本稿では，C. R. Zou 等の非因果性の 2 次元 AR モデル [2] に基づく FRLS [3] を，時間的な因果性を満たす非対称半平面モデルについて拡張し，第 2 節の手法の高速化を行った．また，C. R. Zou 等の手法を拡張する際に改善も行った．C. R. Zou 等の手法では，正規方程式と式 (3) を一つにした拡張正規方程式に基づいて FRLS が導出されており，その際に

- 誤差の 2 乗和の最小値 $J_{min}(k)$ を 1 と仮定することで係数ベクトルの更新式を得ている
- 係数$\hat{a}_{0,0}(k) = 1$ が常に推定される

という問題点があるが，本手法ではこれらの点が以下のように改善されている．

1. $J_{min}(k)$ を 1 と仮定しないため
   - 係数ベクトルの更新式の導出がより一般的
   - $J_{min}(k)$ を逐次求めることが可能
2. $\hat{a}_{0,0}(k)$ が係数更新式中に含まれていないため
   - $\hat{a}_{0,0}(k) \neq 1$ の時のその他の係数の正規化が不要
   - カルマン利得ベクトルの次数が 1 小さい分だけ計算量が減少

4 シミュレーション結果

本節では合成音声に雑音を加えて生成された平面波データに対して実験を行い，提案する手法と 1 次元の適応的手法により得られる推定スペクトルを比較する．

平面波はアレーマイクロホンに正面から入る音声信号を $x_{/i/}$，異なる角度からの有色雑音を $x_{/a/}$ と $x_{/o/}$ ($x_{/i/}$ に対して S/N がそれぞれ 10 dB) とし，白色雑音は $x_{/i/}$ に対して S/N を 20 dB として作成された．音声 /i/ の規範スペクトルと 1 本のマイクロホンで観測された信号に対して 1 次元の適応的手法 [4] を用いて推定されたスペクトル，そして，本手法により推定された 2 次元スペクトルからアレーマイクロホンの正面より観測される信号の推定スペクトルのみを切り出したをものを図 1 に示す．図より，1 次元の手法では音声のスペクトルは雑音の影響を受けて劣化しているが，本手法ではそれよりかなり正確なスペクトルが推定されていることがわかる．

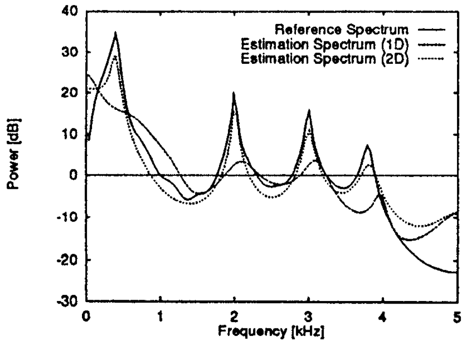

図 1: 規範スペクトルと推定スペクトルとの比較

5 まとめ

本稿ではアレーマイクロホンに異なる角度から入射する音声信号と雑音のスペクトルを分離して推定するため，C. R. Zou 等により提案されている 2 次元非因果性 AR モデルに基づく FRLS を，非対称半平面の AR モデルに基づいて拡張した．また，係数更新式の一般化と計算量の削減の点からの改善を行った．そして，合成音声と雑音からなる合成平面波についてシミュレーションを行い，1 次元の適応的手法では雑音の影響を受けて音声のスペクトル推定精度は劣化するが，本手法では雑音と音声のスペクトルが分離されて，音声のスペクトル推定精度がそれよりかなりよくなることを示した．

今後の課題としては，本手法の有限精度演算の点からの数値安定性を向上させること，係数推定に要するマイクロホン数を減少させること等が挙げられる．

参考文献

[1] 大賀 寿郎，山崎 芳男，金田 豊：“音響システムとディジタル処理”, 電子情報通信学会, 1995

[2] 田口 亮，浜田 望：“多次元ディジタル信号処理における線形予測法”, コンピュートロール, **30**, pp.50–60, Apr., 1990

[3] C. R. Zou, E. I. Plotkin and M. N. S. Swamy : “2-D Fast Kalman Algorithms for Adaptive Parameter Estimation of Nonhomogeneous Gaussian Markov Random Field Model”, IEEE Trans. CAS II, **41**, 10, pp.678–692, Oct., 1994

[4] 堀田 英輔，宮永 喜一，栃内 香次：“時変 ARMA-D モデルを用いた適応的音声分析法”, 信学論 (A), **J75–A**, 8, pp.1324–1332, Aug., 1992