計測自動制御学会東北支部　第281回研究集会（2013.6.21）
資料番号281-7

多重演奏音の分析に関する基礎的検討
A Fundamental study on Analyses of Polyphonic Musical Sound

○工藤憲昌　　田所嘉昭‡
○Norimasa Kudoh　　Yoshiaki Tadokoro‡
八戸高専　‡豊橋技科大 工学部
Hachinohe National College of Tech.　Toyohashi Univ. of Tech.

キーワード：勾配アルゴリズム(gradient learning algorithm)，確率的アルゴリズム(stochastic learning algorithm)，多重演奏音(polyphonic musical sound)

連絡先：〒039-1192　八戸市田面木上野平16-1　八戸高専　電気情報工学科
tel:0178-27-7281，　e-mail:kudohk-e@hachinohe-ct.ac.jp

1. はじめに

音楽演奏におけるピッチ（音高）推定の手法として，音源毎に各単音のスペクトルをテンプレートとしてマッチングを行う方法[1]，[1]を拡張して非線形なDPマッチング手法[2]などが提案されている．これらは，テンプレートに対して，マッチングを行うスペクトルをFFTにより求めている．同一オクターブの周波数が調和関係にないことからスペクトル漏れにより精度が低下することが考えられる．また，近年，検索やデータベース作成を目的として，オーケストラなどの多楽器演奏音から音源分離の検討が進められている[3]-[4]．これらは，楽器の演奏音を，ある状態から他の状態への確率的な遷移と考える隠れマルコフモデル(HMM)に基づいている．状態と推移確率の事前学習が重要であり，その精度により音源分離の性能が決められる．加えて，斬新な音楽の分野の出現もあり，上記の事前学習の結果が大きく異なることもあり，十分な性能を出すまでにはいたっていないようである．

これまで，正弦波の振幅を学習パラメータとするLMS(Least Mean Square)型のフーリエアナライザ[5]-[6]を用いてピッチの変動を伴う楽音（神楽の演奏音）の採譜を行ってきた．スペクトル漏れが少なく時間領域でどんな処理ができるのだろうかと考え検討を行った．本稿では，多楽器音からなる演奏音の採譜の前処理として，それぞれの楽器のスペクトル構造をテンプレートとし，確率的な方法と勾配を用いる方法を組み合わせることで振幅と位相を調整し，音源分離を適応的に行う基礎的な検討を行ったので報告する．

以下に本稿の構成を示す．2.では音源推定の構成を述べる．まず，LMS型フーリエアナライザを用いたスペクトル構造を示し，そのあと，多峰性誤差関数に対応する適応アルゴリズムについて説明する．3.では基本的なシミュレーション結果を示す．4.はまとめである．

2.音源推定の構成

2.1 モデル化

図1に単一楽音の波形の例を示す.図示するように，楽音は，過渡的な状態から定常状態に至り，減衰していく．本稿では，定常状態におけるスペクトル構造を元に適応動作を行うものとする．定常状態における楽音信号$s(n)$を式(1)のようにモデル化する．よく知られているように，楽音はオクターブ間では基本周波数とその整数倍の周波数から構成され

る調和構造となっている．ωは基本角周波数であり，jはωの整数倍を示すものになっている．A_iはi番目の楽器のゲイン，a_{ij}とb_{ij}は定常状態におけるそれぞれ cos 成分，sin 成分の振幅であり，θ_iは位相である．

$$s(n)=\sum_{i=1}^{I} A_i \sum_{j=1}^{J} [a_{ij}\sin(j\omega n - j\theta_i) + b_{ij}\sin(j\omega n - j\theta_i)] \quad (1)$$

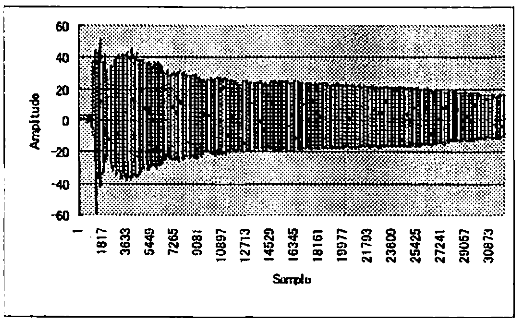
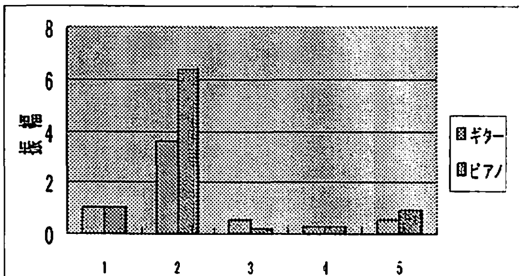


図 1. 単一楽音の波形の例
Fig.1 An example of a single tone wave

オクターブ5のドの定常状態におけるギターとピアノ音のa_{ij}とb_{ij}の値を第5調波成分まで図2に示す．前述したように，同一オクターブの周波数は調和関係になっていない．この場合でもスペクトル漏れが生じないように，2.2 で示す周波数既知を前提とした LMS 型フーリエアナライザで測定したものである．



i		1	2	3	4	5
phase	guitar	0.000	-3.365	-2.340	-2.013	-2.370
	piano	0.000	4.195	-1.241	-1.026	0.676

図 2. オクターブ5のスペクトル構造
Fig.2 Frequency distribution of musical sound at the octave 5

2.2 LMS 型フーリエアナライザ[5]

LMS 型フーリエアナライザ（LMS 法）は，既知の周波数における振幅を逐次推定するアルゴリズムである．この方法では，入力信号$x(n)$とその推定信号$\hat{x}(n)$との推定誤差$e(n)$の最小化を行うことで振幅を推定する．入力信号を式(2)としたとき，推定信号および振幅の更新式はそれぞれ式(3)，(4)となる．ここで，ω_iは既知の角周波数である．

$$x(n)=\sum_{i=1}^{P}\{a_i\cos\omega_i n + b_i\sin\omega_i n\}+\phi(n) \quad (2)$$

$$\hat{x}(n)=\sum_{i=1}^{P}\{\hat{a}_i\cos\omega_i n + \hat{b}_i\sin\omega_i n\} \quad (3)$$

$$\hat{a}_i(n+1)=\hat{a}_i(n)+\mu\cdot e(n)\cos\omega_i n \quad (4\text{-a})$$

$$\hat{b}_i(n+1)=\hat{b}_i(n)+\mu\cdot e(n)\sin\omega_i n \quad (4\text{-b})$$

LMS 法の定常状態における周波数特性を図3に示す．図3のように定常状態における LMS 法の周波数特性は，推定対象の周波数 f_i を中心周波数とする BPF 特性を持つ．適応ループ内に積分操作を盛り込むことにより中心帯域幅を広くすること，つまり，ピッチ（音高）の変動に対応することができる．

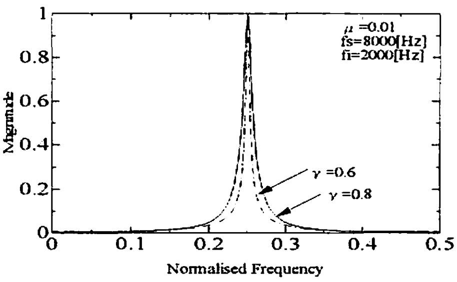


図 3. 定常状態における周波数特性
Fig.3 Frequency characteristics in the steady state

2.3 適応アルゴリズム

2.2 で求めた定常状態におけるスペクトル構造を元にした式(1)の推定値$\hat{s}(n)$は，時刻 n におけるA_iの推定値$\hat{A}_i(n)$，位相θ_iの推定値$\hat{\theta}_i(n)$を用いると，式(5)のようになる．

$$\hat{s}(n)=\sum_{i=1}^{I}\hat{A}_i(n)\sum_{j=1}^{J}[a_{ij}\sin(j\omega n - j\hat{\theta}_i(n)) + b_{ij}\sin(j\omega n - j\hat{\theta}_i(n))] \quad (5)$$

s(n)と$\hat{s}(n)$の誤差 e(n)の2乗値を最小化するように，推定値を学習していく図4の構成を考える．

入力信号式(1)の s(n)に対して，A_iは線形であり，θ_iは非線形である．このため，図5にゲインと位相を変えたときの推定誤差 $e^2(n)$の概形を示すが，単峰性ではなく多峰性になっていることが分かる．

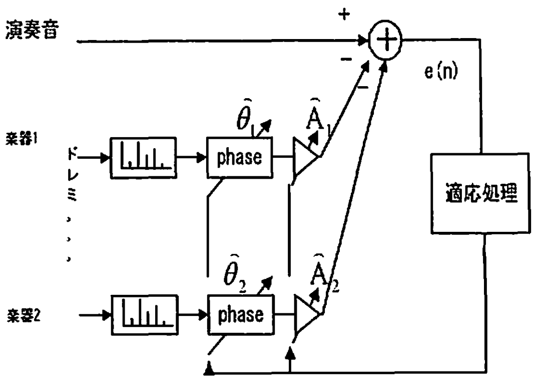


図 4. 音源推定の構成
Fig.4 configuration of adaptive algorithm

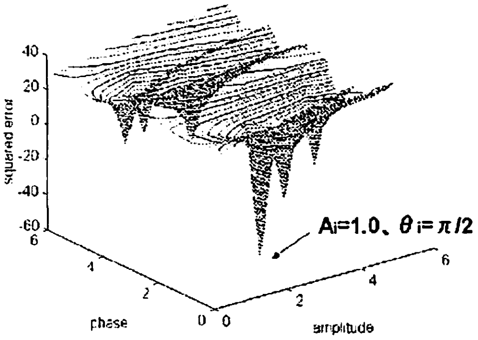


図 5. e(n)の誤差局面
Fig5 Performance surface of e(n)

このように，誤差関数が多峰性のため，勾配法では局所解に陥り最適解に到達できない可能性があり，またランダムな動作をする遺伝的アルゴリズムなどでは収束速度が規定できない．このため，学習の収束速度が速いガウス-ニュウトン法（勾配法）とランダムな動作を行う LRS 法（Linear Random Search）[7] を組み合わせることを考える．

まず，LRS について説明する．説明の簡単化のため，2 つの楽器について考える．適応パラメータのゲインと位相をベクトル化して，式(6)，図 6（3 つ並列に行う場合）のように示す．なお，T は転置を示す

$$\hat{x}(n) = \left(\begin{pmatrix} \hat{A}_1(n) \\ \hat{\theta}_1(n) \end{pmatrix}, \begin{pmatrix} \hat{A}_2(n) \\ \hat{\theta}_2(n) \end{pmatrix} \right)^T \quad (6)$$

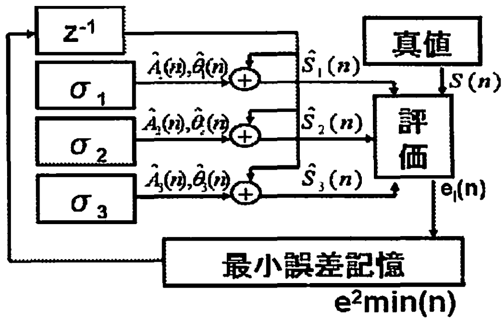


図 6. LRS 法のブロック図
Fig.6 A block diagram of the LRS method

LRS アルゴリズムは，式(7)のようになる．ここで，$\Delta(\sigma)$は，平均ゼロ，分散σ^2のガウス雑音ベクトルである．

$$\hat{x}(n+1) = \hat{x}(n) + \Delta(\sigma) \quad (7)$$

次に，ガウス-ニュウトン法[8]について示す．よく知られているように，関数 f の $f(x_n)=0$ に対して，再帰的に解 x を求めるためには，式(8)を用いる．

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad (8)$$

付録に，詳細を示すように，ゲイン A i と位相 θ i の推定値は 2 階微分（ヘシアン行列）を近似することにより，以下のように更新される．

$$\hat{A}_i(n+1) = \hat{A}_i(n) - P_{gain}(n)\frac{\partial e(n)}{\partial \hat{A}_i(n)} e(n) \quad (9)$$

$$\hat{\theta}_i(n+1) = \hat{\theta}_i(n) - P_{ph}(n)\frac{\partial e(n)}{\partial \hat{\theta}_i(n)} e(n) \quad (10)$$

確率的な LRS と勾配法のガウス-ニュウトン法をその都度更新し，両者のうち誤差の小さい方を推定値とする．

3. 数値例

付録の式(A8)，(A9)の忘却変数 $\lambda=0.95$ として，ギターとピアノの音源を混合して，適切に推定できるか確認する．なお，LRS の分散の比は，1:4:9 とし，LRS の推定値が用いられた場合，勾配法の学習履歴をできるだけ早く低減するよう λ を一時的に小さな値（$\lambda=0.5$）に設定している．サンプリング周波数は 44[kHz]とした．

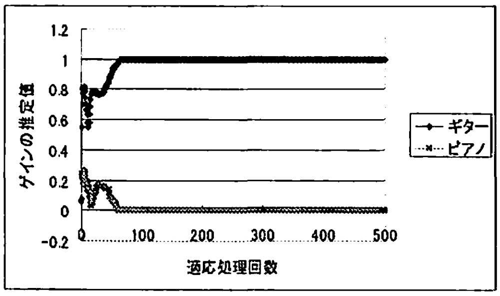


図 7. ギター音の推定結果
Fig.7 Estimated result (guitar)

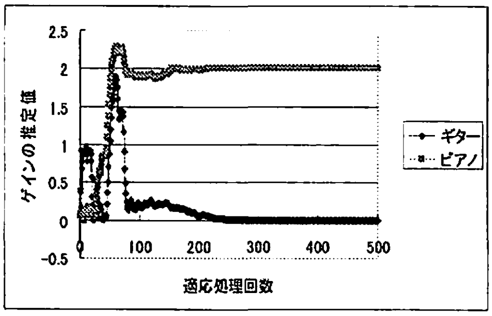


図 8. ピアノ音の推定結果
Fig.8 Estimated result (piano)

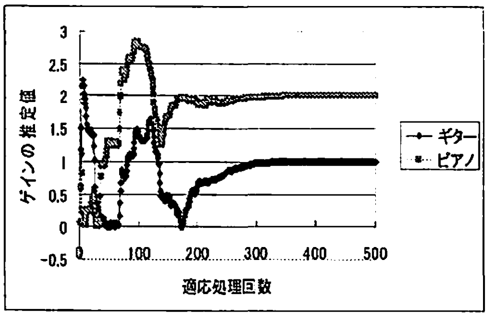


図 9. 合成音の推定結果
Fig.9 Estimated result (guitar and piano)

4. まとめ

それぞれの楽器の定常状態におけるスペクトル構造をテンプレートとし，振幅と位相を LRS アルゴリズムとガウス－ニュウトン法により調整することで，音源分離を適応的に行う基礎的な検討を行った．いくつかの音高に対して，計算機による合成音では，正しく音源を推定し，分離することができた．

今後は，実際の演奏を考慮し，非定常状態期間について，LMS 型フーリエアナライザと組み合わせることを検討する．また，複数楽器間の定常状態のオーバーラップが少ない状況や周波数変動がある場合について検討していく予定である．

5. 参考文献

[1] 中蘆，柏野，田中 “音楽音響信号を対象とする音源分離システム，”情報処理学会技術報告書，SIGMUS-1, pp.1-8

[2]亀岡，篠田，嵯峨山 “スペクトル領域のDPマッチングによる自然楽器演奏の多重音解析，” www/hil.t.u-tokyo.ac.jp/publications/download.php?bib...pdf

[3]Tetsuro Kitahara et al. : 「Instrogram: Probabilistic Instrument Existence for Polyphonic Music」, ISPJ Journal, pp.214-226,Jan. 2007

[4]Essid S. Richard et al.: 「Instrument Recognition in Polyphonic Music」,Proc. ICASSP, Vol.Ⅲ, pp.245-248, 2005

[5]N.Kudoh, Y.Tadokoro 「Performance Analysis of an LMS based Fourier Analyzer for Sinusoidal signals with Time-varying Amplitude」 CD-ROM Proceedings of TENCON'03(IEEE Region 10 Technical Conference), pp.1-5, Bangalore, India(2003年10月)

[6]渡辺，工藤，田所「LMS型フーリエアナライザを用いた神楽音の分析Ⅱ」計測自動制御学会第258回研究集会　258-18, pp.1-5(2010年6月)

[7]谷萩著：「ディジタル信号処理の理論-3 推定・適応信号処理」コロナ社

[8]L.Ljung : System Identification: Theory for the user, Prentice Hall,1987

付録

ガウス－ニュウトン法では，式(A1)の評価関数ξの適応パラメータに関する1階，2階微分が必要になる．

$$f(x) = \frac{de^2(n)}{d\hat{x}} = \frac{d\xi(x)}{d\hat{x}} \qquad (A1)$$

ここで，$\hat{x}(n)$は式(A2)に示す適応パラメータのベクトルである．ここで，T は転置を示す。

$$\hat{x}(n) = \left(\begin{pmatrix} \hat{A}_1(n) \\ \hat{\theta}_1(n) \end{pmatrix}, \begin{pmatrix} \hat{A}_2(n) \\ \hat{\theta}_2(n) \end{pmatrix} \right)^T \qquad (A2)$$

1 階微分は式(A3)で与えられ，幅と位相の勾配成分（e(n)に関する微分）について式(A4)に示す．

$$\xi'(\hat{x}(n)) = \frac{1}{2}\frac{de^2(n)}{d\hat{x}(n)} = e(n)\frac{de(n)}{d\hat{x}(n)} \qquad \text{(A3)}$$

$$\frac{\partial e(n)}{\partial \hat{A}_i(n)} = \psi_{Ai} = -\sum_{j=1}^{J}[a_{ij}\cos(j\omega n - j\hat{\theta}_i(n)) + b_{ij}\sin(j\omega n - j\hat{\theta}_i(n))] \qquad \text{(A4-1)}$$

$$\frac{\partial e(n)}{\partial \hat{\theta}_i(n)} = \psi_{\theta i} = -\hat{A}_i(n)\sum_{j=1}^{J}[j\cdot a_{ij}\sin(j\omega n - j\hat{\theta}_i(n)) - j\cdot b_{ij}\cos(j\omega n - j\hat{\theta}_i(n))] \qquad \text{(A4-2)}$$

2階微分（ヘシアン行列）は，適応パラメータの更新に対する誤差の変化が小さいとして，式(A5)に示すように近似している．

$$\xi''(\hat{x}(n)) = \frac{de(n)}{d\hat{x}(n)} \times \left(\frac{de(n)}{d\hat{x}(n)}\right)^T + e(n)\frac{d^2e(n)}{d\hat{x}(n)^2} \approx \frac{de(n)}{d\hat{x}(n)}\left(\frac{de(n)}{d\hat{x}(n)}\right)^T \qquad \text{(A5)}$$

ξの 2 階微分は式(5)に示すように分母に用いるため，次式に示す逆行列の補助定理を用いて変換する．

$$[A+BCD]^{-1} = A^{-1} - A^{-1}B[DA^{-1}B + C^{-1}]DA^{-1} \qquad \text{(A6)}$$

Aをξの2階微分，$B = D^T = \psi_{Ai}$，Cを単位行列，$P_{gain}(n)$をξの2階微分の逆行列とすると，$P_{gain}(n)$は以下のように計算される．なお，Iは単位行列である．

$$P_{gain}(n) = \left[I - \frac{\left(\frac{de(n)}{d\hat{x}(n)}\right)\cdot\left(\frac{de(n)}{d\hat{x}(n)}\right)^T \cdot P_{gain}(n-1)}{\left(\frac{de(n)}{d\hat{x}(n)}\right)^T \cdot P_{gain}(n-1)\cdot\left(\frac{de(n)}{d\hat{x}(n)}\right) + 1}\right] \times P_{gain}(n-1) \qquad \text{(A7)}$$

実際には，雑音などの影響を考慮し，忘却変数λ ($0<\lambda<1$)を導入して，振幅(添え字gain)，位相(添え字ph)に対して式(A8)，(A9)のように求められる．

$$P_{gain}(n) = \frac{1}{\lambda}\left[I - \frac{\left(\frac{de(n)}{d\hat{x}(n)}\right)\cdot\left(\frac{de(n)}{d\hat{x}(n)}\right)^T \cdot P_{gain}(n-1)}{\left(\frac{de(n)}{d\hat{x}(n)}\right)^T \cdot P_{gain}(n-1)\cdot\left(\frac{de(n)}{d\hat{x}(n)}\right) + \lambda}\right] \times P_{gain}(n-1) \qquad \text{(A8)}$$

$$P_{ph}(n) = \frac{1}{\lambda}\left[I - \frac{\left(\frac{de(n)}{d\hat{x}(n)}\right)\cdot\left(\frac{de(n)}{d\hat{x}(n)}\right)^T \cdot P_{ph}(n-1)}{\left(\frac{de(n)}{d\hat{x}(n)}\right)^T \cdot P_{ph}(n-1)\cdot\left(\frac{de(n)}{d\hat{x}(n)}\right) + \lambda}\right] \times P_{ph}(n-1) \qquad \text{(A9)}$$

従って，振幅と位相の更新式は以下のようになる．

$$\hat{A}_i(n+1) = \hat{A}_i(n) - P_{gain}(n)\frac{\partial e(n)}{\partial \hat{A}_i(n)}e(n) \qquad \text{(A10)}$$

$$\hat{\theta}_i(n+1) = \hat{\theta}_i(n) - P_{ph}(n)\frac{\partial e(n)}{\partial \hat{\theta}_i(n)}e(n) \qquad \text{(A11)}$$