非負値行列因子分解に基づく欠損データ補間による声道スペクトル推定法の検討*

◎中村友彦（東大院・情報理工），亀岡弘和（東大院・情報理工，NTT・CS 研）

1 はじめに

音声合成や音声変換をはじめ音声処理全般において，音声信号から声道スペクトルを推定する技術は多くの場面で用いられている。短区間ごとの音声信号を周期デルタ関数（パルス列）を入力とした線形時不変系の出力としてモデル化できるとすれば，この線形系の入力とインパルス応答がぞれぞれ声帯音源信号と声道特性に対応する。この仮定は，周波数領域では周期デルタ関数で表される声帯音源スペクトルと声道スペクトルの積で音声スペクトルが表されることに相当する。従って，音声スペクトルは声道スペクトルを周期的に（基本周波数間隔で）サンプリングしたもの，と見なすことができる。

この観点に基づいて音声スペクトルから声道スペクトルを推定する方法がこれまで多数提案されている。代表的な方法の一つとして広く知られる “STRAIGHT [1]” は，音声信号を基本周期の幅で切り出し，その切り出し信号のスペクトルを声道スペクトルの推定値とする方法である。これは周波数領域では，各調波成分のピークを sinc 補間したものを声道スペクトルと見なしていることに相当する。しかしこの方法によって得られる声道スペクトル推定値は，定常な音声が対象であっても切り出しフレームのオフセットに依存して周期的に時間変化することが知られる。これは各調波成分が互いに干渉し合うからである。この変動成分は周期信号に対する有限窓を用いた周波数分析により不可避的に生じるものであり，声道スペクトル推定値に本来含めるべきものではない。そこで，[2] ではこの変動成分を除くよう改良された手法が提案されている。

前述のとおり音声スペクトルは声道スペクトルを基本周波数 (F_0) 間隔でサンプリングしたものと見なせるため，音声の F_0 が高いときほど声道スペクトル推定の手がかりは少なくなる。このことは，フレームごとに独立な処理に本質的な限界があることを示唆している。一方で，音声信号には同一の音素が繰り返し出現するため，類似した声道スペクトルが複数の異なる時刻で現れることも手がかりとなる．複数のフレームが共通の声道スペクトルを持つと仮定できそれらのフレームで F_0 が異なれば，実際に観測可能な声道スペクトルのサンプル点が単一のフレームの場合よりも増加するので，スペクトル推定精度が向上すると考えられる．このような考えに基づき，同時に収録された調音運動データを用いて複数フレームから声道スペクトルを推定する手法 [3] が提案されている．また，同様の手法として，因子分析トラジェクトリ隠れマルコフモデルによる声道スペクトル推定法が提案されている [5]. この手法では，音声の各フレームに付与されているコンテキストラベルを用い，同一のコンテキストが付与された複数のフレームにおける調波成分の情報や動的特徴量を手がかりにすることでスペクトル成分を推定する．

以上のように，[5] の手法では複数フレームにおける調波成分の情報を手がかりに高精度に声道スペクトルを推定することが可能であるが，音声データに対するコンテキストラベルの付与には膨大な労力を要する。そこで本稿では，声道スペクトログラムが低ランクな非負値行列で近似できるという仮定を元に，コンテキストラベルを用いなくとも複数フレームにおける調波成分の情報を手がかりに高精度に声道スペクトルを推定することができる手法を提案する。

2 欠損データ補間による声道スペクトル推定の定式化

2.1 声道スペクトログラムモデルの設計

時刻インデックスを $t = 0, \cdots, T-1$ とし，周波数インデックスと対応する正規化角周波数をそれぞれ $k = 0, \cdots, K-1$, ω_k と表す．前節で述べた通り，声道スペクトログラムを低ランクな非負値行列とみなす．R 個の滑らかなスペクトルパターンを列方向に並べた非負値行列 $H = (H_{k,r})_{k,r}$ と，各スペクトルパターンの非負値の重み $U = (U_{r,t})_{r,t}$ を用いて，声道スペクトログラム $X_{k,t}$ は，

$$X_{k,t} = \sum_{r=0}^{R-1} H_{k,r} U_{r,t} \tag{1}$$

と表現できる．r はスペクトルパターンのインデックスである．

周波数方向に滑らかかつ非負となる $H_{k,r}$ は様々に設計可能であるが，本稿では 2 種類の $H_{k,r}$ を提案す

---

* Vocal Tract Spectrum Estimation Based on Missing Data Interpolation Using Non-Negative Matrix Factorization by NAKAMURA, Tomohiko (The University of Tokyo) and KAMEOKA, Hirokazu (The University of Tokyo / Nippon Telegraph and Telephone Corporation)

る．1 つ目の設計では，[6] と同様に声道スペクトルを混合正規分布（GMM）型の関数

$$H_{k,r}^{(\mathrm{GMM})} = \sum_n W_{r,n} G_n(\omega_k) \tag{2}$$

$$G_n(\omega) = \frac{1}{\sqrt{2\pi\nu_n^2}} e^{-\frac{(h(\omega)-\rho n)^2}{2\nu_n^2}} \tag{3}$$

で記述する．ここで，ω は正規化角周波数，N は混合数，$n = 0, \cdots, N-1$ はそのインデックスである．$h(\omega)$ は周波数ワーピング関数であり，$h(\omega) = \omega$ とすれば $G_n(\omega)$ は線形周波数領域で平均 ρ_n，分散 ν_n^2 の正規分布と同形の関数となる．本稿では，メル周波数領域で滑らかな声道スペクトルとなるように以下のように $h(\omega)$ を設計した．

$$h(\omega) = \frac{\ln(\alpha\omega + 1)}{\ln(\alpha\pi + 1)}, \quad \alpha = \frac{1}{2\pi} \times \frac{f_s}{625}. \tag{4}$$

ここで，f_s はサンプリング周波数である．(4) 式により $[0, \pi]$ の正規化周波数は $[0, 1]$ にマッピングされる．$W = (W_{r,n})_{r,n} \geq 0$ は各正規分布の重みであり，H, U に関するスケールの任意性を解消するため各 r に関し $\sum_n W_{r,n} = 1$ とする．

2 つ目の設計では，ソースフィルタモデルでよく用いられる全極フィルタを利用する．P 次の全極フィルタの係数 $\boldsymbol{a}_r := [a_{r,0}, a_{r,1}, \cdots, a_{r,P}]^\top$ を用いて，全極フィルタの振幅スペクトルは

$$H_{k,r}^{(\mathrm{AR})} = \frac{1}{(\boldsymbol{a}_r^\top Q(\omega_k) \boldsymbol{a}_r)^{1/2}} \tag{5}$$

と表せる．ここで，$Q(\omega)$ は (p, q) 成分が $\cos(\omega(p-q))$ で表される $(P+1) \times (P+1)$ の Toeplitz 行列である．

2.2 非負値行列因子分解による定式化

STRAIGHT により推定されたスペクトログラム $Y = (Y_{k,t})_{k,t}$ が与えられたとき，声道スペクトログラム推定問題は $Y_{k,t}$ と $X_{k,t}$ の距離 $D_*(Y_{k,t}; X_{k,t})$ を用いて，

$$\min_\Theta \mathcal{L}_*(\Theta) = \sum_{k,t} Z_{k,t} D_*(Y_{k,t}; X_{k,t}) \tag{6}$$

と定式化できる．ここで，Θ はパラメータ集合であり，$H_{k,r}$ として $H_{k,r}^{(\mathrm{GMM})}$ を用いた場合は $\Theta = \{W, U\}$，$H_{k,r}^{(\mathrm{AR})}$ を用いた場合は $\Theta = \{\{\boldsymbol{a}_r\}_r, U\}$ である．以後簡単のため，$H_{k,r}^{(\mathrm{GMM})}$ を用いた場合の NMF を GMM-NMF，$H_{k,r}^{(\mathrm{AR})}$ を用いた場合は AR-NMF と呼ぶ．$Z_{k,t} \in [0, 1]$ は各時間周波数成分の信頼度を表すパラメータである．$Z_{k,t} = 0$ であれば $Y_{k,t}$ に対するコストは全く考慮せず，$Z_{k,t}$ が大きい時間周波数ビンほど重視される．信頼度を決定するために，STRAIGHT 分析で得られる非周期性指標 $A_{k,t} \in [0, 1]$ が利用できる．非周期性指標は各時間周波数ビンに含まれる非周期成分の割合であるため，各 k, r に関して $Z_{k,t} = 1 - A_{k,t}$ とすれば周期性成分を重視した推定が可能である．

NMF で広く用いられてる距離 D_* として，一般化 Kullback-Leibler（KL）ダイバージェンス D_{GKL}，2 乗距離 D_{EU} が挙げられる．それぞれの距離またはダイバージェンスを用いたときの目的関数 $\mathcal{L}_*(\Theta)$ は，

$$\begin{aligned}\mathcal{L}_{\mathrm{GKL}}(\Theta) = \sum_{k,t} Z_{k,t} \,(&Y_{k,t} \ln Y_{k,t} - Y_{k,t} \ln X_{k,t} \\ &- Y_{k,t} + X_{k,t})\end{aligned} \tag{7}$$

$$\mathcal{L}_{\mathrm{EU}}(\Theta) = \sum_{k,t} Z_{k,t} (Y_{k,t} - X_{k,t})^2 \tag{8}$$

と書ける．

3 補助関数法によるパラメータ推定アルゴリズム

3.1 GMM-NMF に対する反復アルゴリズムの導出

本節では GMM-NMF に対するパラメータ推定アルゴリズムを導出し，AR-NMF に関しては次節で扱う．まず，(7) 式で定義される目的関数を考えよう．(7) 式の括弧内の第 2 項は対数関数の中に和を含んでおり，直接最適化問題を解くことは難しい．しかし，多くの NMF を用いた研究で使われているように，補助関数法 [7,8] と呼ばれる最適化原理によって反復的に局所最適解を得ることができる．補助関数法では，パラメータ Θ の目的関数 $\mathcal{L}(\Theta)$ に対して補助変数 λ を導入し，$\mathcal{L}(\Theta) = \min_\lambda \mathcal{L}^+(\Theta, \lambda)$ を満たす上界 $\mathcal{L}^+(\Theta, \lambda)$（補助関数）を導出する．$\mathcal{L}^+(\Theta, \lambda)$ を Θ, λ に関して交互に最小化することによって，$\mathcal{L}(\Theta)$ を広義単調減少させることができる．

対数関数は凹関数なので，(7) 式の第 2 項の上界は Jensen の不等式を用いて以下のように導出できる．

$$\begin{aligned}&- Z_{k,t} Y_{k,t} \ln \left(\sum_{r,n} W_{r,n} G_n(\omega_k) U_{r,t} \right) \\ &\leq - Z_{k,t} Y_{k,t} \sum_{r,n} \lambda_{k,t,r,n} \ln \left(\frac{W_{r,n} G_n(\omega_k) U_{r,t}}{\lambda_{k,t,r,n}} \right)\end{aligned} \tag{9}$$

ここで，$\lambda_{k,t,r,n} \geq 0$ は補助変数であり，各 k, t に関し $\sum_{r,n} \lambda_{k,t,r,n} = 1$ を満たす．等式成立条件は，

$$\lambda_{k,t,r,n} = \frac{W_{r,n} G_n(\omega_k) U_{r,t}}{X_{k,t}} \tag{10}$$

である．したがって，$\mathcal{L}_{\mathrm{GKL}}(\Theta)$ に対する補助関数は

$$\begin{aligned}&\mathcal{L}^{+}_{\mathrm{GKL,GMM}}(\Theta,\lambda)\\&=\sum_{k,t} Z_{k,t}Y_{k,t}\left(-\sum_{r,n}\lambda_{k,t,r,n}\ln\frac{W_{r,n}G_n(\omega_k)U_{r,t}}{\lambda_{k,t,r}}\right.\\&\quad\left.+\ln Y_{k,t}\right)+\sum_{k,t} Z_{k,t}(X_{k,t}-Y_{k,t}) \qquad (11)\end{aligned}$$

と導出できる．ここで，$\lambda := \{\lambda_{k,t,r,n}\}$ とした．補助関数の $W_{r,n}, U_{r,t}$ に関する偏微分が 0 となる値を求め (10) 式を代入することにより，閉形式の更新式が以下のように得られる．

$$W_{r,n} \propto W_{r,n}\frac{\sum_{k,t} Z_{k,t}\frac{Y_{k,t}}{X_{k,t}}G_n(\omega_k)U_{r,t}}{\sum_{k,t} Z_{k,t}G_n(\omega_k)U_{r,t}} \qquad (12)$$

$$U_{r,t} \leftarrow U_{r,t}\frac{\sum_{k} Z_{k,t}\frac{Y_{k,t}}{X_{k,t}}H^{(\mathrm{GMM})}_{k,r}}{\sum_{k} Z_{k,t}H^{(\mathrm{GMM})}_{k,r}} \qquad (13)$$

両更新式は全て非負値の項同士の積として計算されるため，初期値を非負値にすれば W, U の非負値性は自然と保たれる．また，(4) 式では α を定数としたが，

2 乗距離を用いた場合の目的関数に関しても考える．(8) 式は括弧内の第 2 項に和を含んでおり直接最適化問題を解くことは難しいが，2 次関数は凸関数なので Jensen の不等式を用いて前述の場合と同様に補助関数を設計できる．補助関数の導出について詳細は紙面の都合上省くが，更新式は以下のように得られる．

$$W_{r,n} \leftarrow W_{r,n}\frac{\sum_{k,t} Z_{k,t}Y_{k,t}G_n(\omega_k)U_{r,t}}{\sum_{k,t} Z_{k,t}X_{k,t}G_n(\omega_k)U_{r,t}} \qquad (14)$$

$$U_{r,t} \leftarrow U_{r,t}\frac{\sum_{k} Z_{k,t}Y_{k,t}H^{(\mathrm{GMM})}_{k,r}}{\sum_{k} Z_{k,t}X_{k,t}H^{(\mathrm{GMM})}_{k,r}} \qquad (15)$$

3.2 AR-NMF に対する反復アルゴリズムの導出

AR-NMF に関しても同様に補助関数法を用いて閉形式の更新式を導出できる．前節と同様に，まず一般化 KL ダイバージェンスの場合を考える．Jensen の不等式を用いて補助関数は，各 k,t に関し $\sum_r \xi_{k,t,r} = 1$ を満たす非負の補助変数 $\xi = \{\xi_{k,t,r}\}_{k,t,r}$ を導入し，

$$\begin{aligned}&\mathcal{L}^{+}_{\mathrm{GKL,AR}}(\Theta,\xi)\\&=\sum_{k,t} Z_{k,t}Y_{k,t}\left(\ln Y_{k,t}-\sum_{r}\xi_{k,t,r}\ln\frac{H^{(\mathrm{AR})}_{k,r}U_{r,t}}{\xi_{k,t,r}}\right)\\&\quad+\sum_{k,t} Z_{k,t}(X_{k,t}-Y_{k,t}) \qquad (16)\end{aligned}$$

と定義できる．等式成立条件は $\xi_{k,t,r} = H^{(\mathrm{AR})}_{k,r}U_{r,t}/X_{k,t}$ である．$U_{r,t}$ の更新式は，(13) 式の $H^{(\mathrm{GMM})}_{k,r}$ を $H^{(\mathrm{AR})}_{k,r}$ に置換したものと同一である．

一方で，$\boldsymbol{a}_r$ の更新には乗法更新型アルゴリズムを利用できる [9,10]. $\mathcal{L}_{\mathrm{GKL}}(\Theta)$ の $\boldsymbol{a}_r$ に関する偏微分は,

$$\frac{\partial\mathcal{L}_{\mathrm{GKL}}}{\partial\boldsymbol{a}_r}(\Theta) = (\Psi^{+}_r - \Psi^{-}_r)\boldsymbol{a}_r \qquad (17)$$

$$\Psi^{+}_r = \sum_{k,t} Z_{k,t}\frac{Y_{k,t}}{X_{k,t}}(H^{(\mathrm{AR})}_{k,r})^3U_{r,t}Q(\omega_k) \qquad (18)$$

$$\Psi^{-}_r = \sum_{k,t} Z_{k,t}(H^{(\mathrm{AR})}_{k,r})^3U_{r,t} \qquad (19)$$

と書ける．上式の括弧内の第 1 項と第 2 項はどちらも正定値行列であり，以下の乗法更新則を用いることで目的関数 $\mathcal{L}_{\mathrm{GKL}}(\Theta,\xi)$ を広義単調減少させる事ができる．

$$\boldsymbol{a}_r \leftarrow (\Psi^{+}_r)^{-1}\Psi^{-}_r\boldsymbol{a}_r. \qquad (20)$$

詳細は省くが，2 乗距離の場合においても同様に更新式を導出することができる．$U_{r,t}$ の更新式は，(15) 式の $H^{(\mathrm{GMM})}_{k,r}$ を $H^{(\mathrm{AR})}_{k,r}$ に置換したものと同一であり，$\boldsymbol{a}_r$ の更新式は (20) 式と同様に以下で与えられる．

$$\begin{aligned}\boldsymbol{a}_r \leftarrow &\left(\sum_{k,t} Z_{k,t}Y_{k,t}(H^{(\mathrm{AR})}_{k,r})^3U_{r,t}Q(\omega_k)\right)^{-1}\\&\times\left(\sum_{k,t} Z_{k,t}X_{k,t}(H^{(\mathrm{AR})}_{k,r})^3U_{r,t}Q(\omega_k)\right)\boldsymbol{a}_r\end{aligned} \qquad (21)$$

4 声道スペクトル推定精度評価実験

4.1 実験条件

[5] に倣い，音声信号の分析再合成を通して提案法の声道スペクトルの推定性能評価実験を行った．ATR デジタル音声データベース [11] の A セットから日本人女性話者 1 名による 20 文の音声信号（サンプリング周波数 16 kHz）を STRAIGHT で分析し，F_0，声道スペクトル，非周期性指標を抽出した．ここで得られたスペクトルを正解の声道スペクトルとみなす．正解のスペクトルと $2^{-1.0}, 2^{-0.5}, 2^{0.0}, 2^{0.5}, 2^{1.0}, 2^{1.5}$ 倍した F_0 を用いて，音声信号をそれぞれ STRAIGHT で再合成した．

再合成信号から STRAIGHT または提案法でスペクトルを推定し，スペクトル推定値と正解の声道スペクトルとのメルケプストラム歪み（1 次から 24 次のメルケプストラム係数を用いて計算）を用いて性能を比較した．STRAIGHT 分析ではフレームシフトを 5 ms（$T = 81761$），スペクトルの次元を $K = 513$ とした．GMM-NMF の方が AR-NMF に比べ実装が容易なため AR-NMF の性能評価は今後の課題とし，本実験では一般化 KL ダイバージェンスを用いた

Table 1　一般化 KL ダイバージェンス基準の GMM-NMF（GKL）と 2 乗距離基準の GMM-NMF（EU），STRAIGHT 分析による，F_0 を x 倍された再合成音声信号に対するメルケプストラム歪みの平均値と標準偏差 [dB]．括弧内の値は非周期性指標を用いなかった場合の結果である．

x	GKL	EU	STRAIGHT
$2^{-1.0}$	$\mathbf{3.94 \pm 1.54}(4.01 \pm 1.64)$	$4.33 \pm 1.33(4.49 \pm 1.31)$	4.26 ± 1.88
$2^{-0.5}$	$\mathbf{3.77 \pm 1.53}(3.96 \pm 1.64)$	$4.48 \pm 1.47(4.39 \pm 1.32)$	4.23 ± 1.87
$2^{0.0}$	$\mathbf{3.81 \pm 1.49}(3.95 \pm 1.62)$	$4.31 \pm 1.27(4.46 \pm 1.47)$	4.24 ± 1.85
$2^{0.5}$	$\mathbf{4.36 \pm 1.17}(4.50 \pm 1.23)$	$4.73 \pm 1.28(4.60 \pm 1.46)$	4.80 ± 1.38
$2^{1.0}$	$\mathbf{5.32 \pm 1.56}(5.37 \pm 1.56)$	$5.46 \pm 1.68(5.40 \pm 5.56)$	5.56 ± 1.54
$2^{1.5}$	$\mathbf{5.88 \pm 2.31}(5.97 \pm 2.28)$	$5.74 \pm 2.50(5.90 \pm 2.34)$	6.20 ± 2.21

GMM-NMF（GKL-GMM-NMF）と 2 乗距離を用いた GMM-NMF（EU-GMM-NMF）の 2 種類の手法を提案法として用いた．$Y_{k,t}$ として STRAIGHT で推定された再合成信号のスペクトルを用い，F_0 の各定数倍ごとに 20 発話の $Y_{k,t}$ を同時に用いて W, U を推定した．$R = 90, N = 100$ とし，$n = 0, \cdots, N-1$ に対して $\rho_n = n/(N-1)$,$\nu_n = 1/(N-1)$ とした．W, U は非負の乱数で初期化し，推定アルゴリズムの反復回数は 100 回とした．

4.2　実験結果

提案法と STRAIGHT による推定結果のメルケプストラム歪みを表 1 に示す．F_0 の倍率が高くなるほど観測できる調波成分も少なくなるため，当該フレーム以外の調波成分を利用することによる効果が現れるはずである．実際に，GKL-GMM-NMF の方が STRAIGHT に比べ F_0 が高くなるほどメルケプストラム歪みが少なく，当該フレーム以外の調波成分が声道スペクトル推定に有効であることが確認できる．EU-GMM-NMF も F_0 が高くなるほど STRAIGHT に比べメルケプストラム歪みが小さくなったが，GKL-GMM-NMF に比べると平均的にメルケプストラム歪みが大きく GKL-GMM-NMF の方が声道スペクトル推定により適していると考えられる．

非周期性指標を全ての時間周波数ビンで一様（全ての k, t に関して $Z_{k,t} = 1$）とした提案法での実験も行った．表 1 の括弧内の数値が非周期性指標なしでの提案法のメルケプストラム歪みである．GKL-GMM-NMF では，非周期性指標が性能向上に寄与することが確認できる．また，非周期性指標なしでも STRAIGHT に比べメルケプストラム歪みが少なく，声道スペクトログラムの低ランク仮定が声道スペクトル推定に有用であることが示唆された．

5　結論

本稿では，声道スペクトルの局所的な手がかりと大局的な手がかりを利用した声道スペクトル推定法を提案した．声道スペクトログラムを低ランクな非負値行列とみなし，声道スペクトル推定を欠損データのある NMF として定式化した．滑らかなスペクトルパターンとして，GMM と同形の関数と全極フィルタに基づくモデルをそれぞれ提案し効率的なパラメータ推定アルゴリズムを提案した．評価実験により，提案法の 1 つが STRAIGHT に比べメルケプストラム歪みを少なく声道スペクトルを推定できることを確認した．今後は，男性話者や他の女性話者での大規模実験や AR-NMF の性能評価，他の距離関数の検討が課題である．

謝辞　本研究は JSPS 科研費 26730100，15J0992 の助成を受けたものです．

参考文献

[1] H. Kawahara *et al.*, *Speech Commun.*, 27, pp. 187–207, 1999.

[2] H. Kawahara *et al.*, *Proc. ICASSP*, pp. 3933–3936, 2008.

[3] Y. Shiga *et al.*, *Proc. EUROSPEECH*, pp. 1737-1740, 2003.

[4] T. Nakamura *et al.*, *Proc. ISMIR*, pp. 623–628, 2014.

[5] T. Toda, *Proc. ICASSP*, pp. 3925–3928, 2008.

[6] H. Kameoka *et al.*, *IEEE Trans. ASLP*, 18 (6), pp. 1507–1516, 2010.

[7] J. M. Ortega *et al.*, *SIAM*, 1970.

[8] D. R. Hunter, *J. Comp. Graph. Stat.*, 9 (1), pp. 60–77, 2000.

[9] A. El-Jaroudi *et al.*, *IEEE Trans. SP*, 39 (2), 1991.

[10] R. Badeau *et al.*, *IEEE Trans. Neural. Netw.*, 21 (12), pp. 1869–1881, 2010.

[11] A. Kurematsu *et al.*, *Speech Commun.*, 9 (4), pp. 357–363, 1990.