社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

非負値行列因子分解を用いた欠損データ補間による超解像声道スペクトル推定

中村　友彦†　　亀岡　弘和†,††

† 東京大学　大学院情報理工学系研究科　〒 113–0023 東京都文京区本郷 7–3–1
†† 日本電信電話株式会社　NTT コミュニケーション科学基礎研究所
〒 243–0198 神奈川県厚木市森の里若宮 3–1
E-mail: †Tomohiko_Nakamura@ipc.i.u-tokyo.ac.jp, ††kameoka.hirokazu@lab.ntt.co.jp

あらまし　本報告では，音声信号から声道スペクトルを推定する問題を扱う．声道スペクトルは基本周波数（F_0）間隔で周期的にサンプリングしたものと見なせるため，音声の F_0 が高いほど声道スペクトル推定の手がかりは少なくなる．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．そのため，F_0 の異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れば，複数のフレームの情報を用いることで声道スペクトル推定精度が向上するはずである．そこで，声道スペクトログラムが低ランクな非負値行列で表現できるという仮定を元に，声道スペクトル推定問題を欠損データのある非負値行列因子分解として定式化し，効率的なパラメータ推定アルゴリズムを導出する．評価実験により提案法の有効性を確認した．
キーワード　声道スペクトル推定，非負値行列因子分解，補助関数法

Super-Resolution Vocal Tract Spectrum Estimation with Missing Data Imputation Using Non-Negative Matrix Factorization

Tomohiko NAKAMURA† and Hirokazu KAMEOKA†,††

† Graduate School of Information Science and Technology, the University of Tokyo
7–3–1, Hongo, Bunkyo-ku, Tokyo, 113–0023 Japan
†† NTT Communicatin Science Laboratories, Nippon Telegraph and Telephone Corporation
3-1, Morinosato Wakamiya, Atsugi, Kanagawa, 243–0198 Japan.
E-mail: †Tomohiko_Nakamura@ipc.i.u-tokyo.ac.jp, ††kameoka.hirokazu@lab.ntt.co.jp

Abstract　This report addresses the problem of estimating vocal tract spectra from speech signals. Spectra of speech signals can be interpreted as vocal tract spectra sampled with a sampling period of F_0, and the higher F_0s are, the less observable harmonic components becomes. On the other hand, similar vocal tract spectra appear repeatedly due to multiple appearance of phonemes in speech signals. If we can assume that frames at which F_0s are different have common vocal tract spectra, the accuracy of vocal tract estimation may improve by making effective use of harmonic components in multiple frames. On the basis of the idea, with an assumption that vocal tract spectrograms can be represented by low-rank non-negative matrices, we formulate the problem of vocal tract estimation as non-negative matrix factorizations for missing data and derive an efficient parameter estimation algorithm. Experimental evaluation shows the effectiveness of the proposed method.
Key words　Vocal Tract Spectrum Estimation, Non-Negative Matrix Factorization, Auxiliary Function Approach

1. はじめに

音声合成や音声変換をはじめとする音声処理において，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を入力とした線形時不変システムの出力としてモデル化できると仮定すれば，このシステム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，Fourier 変換領域では

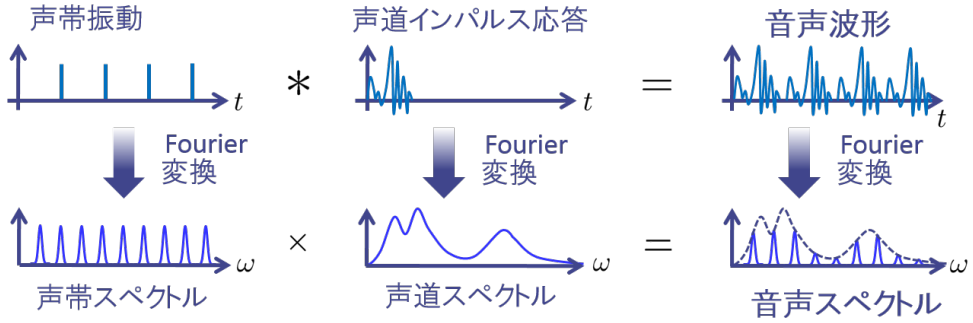


図 1　ソースフィルタモデルによる音声スペクトルの生成過程.

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．したがって，音声スペクトルは声道スペクトルを基本周波数（F_0）間隔で周期的にサンプリングしたものと見なせる（図 1）．

この観点に基づいて音声信号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，切り出された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，各調波成分が干渉しあうため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 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 が高ければ高いほど声道スペクトル推定の手がかりは少なくなる．このため，フレームごとの独立な処理には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．F_0 の異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れば，実際に観測できる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し，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 [4]．この手法では，初期値を得る際に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，[4] 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用いて，コンテキストラベルなしに複数フレームにおける調波成分の情報を手がかりに声道スペクトルを推定できる手法を提案する．

2.　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

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 $k = 0, \cdots, K-1$, $\omega_k \geqq 0$ と表す．声道スペクトログラムが低ランクな非負値行列で表現できると仮定すれば，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が各行に並んだ非負値行列 $U = (U_{r,t})_{r,t}$ の積で，声道スペクトログラム $X_{k,t}$ が

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

と表現できる．ここで，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 $H_{k,r}$ の設計方法は様々なものが考えられるが，本稿では 2 種類の $H_{k,r}$ を提案する．1 つ目の設計では，[5] と同様に声道スペクトルを混合正規分布（GMM）型の関数

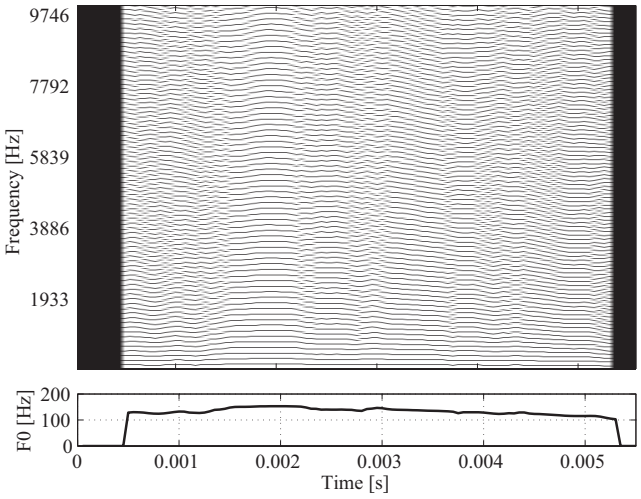
$$H_{k,r}^{(\mathrm{GMM})} = \sum_n W_{r,n} G(\omega_k; \rho_{r,n}, \nu_{r,n}^2) \tag{2}$$

$$G(\omega; \rho_{r,n}, \nu_{r,n}^2) = \frac{1}{\sqrt{2\pi\nu_{r,n}^2}} e^{-\frac{(h(\omega)-\rho_{r,n})^2}{2\nu_{r,n}^2}} \tag{3}$$

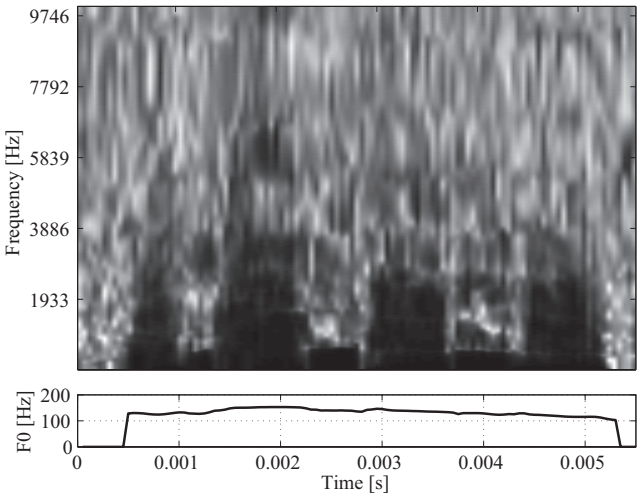
で記述する．ここで，ω は正規化角周波数，N は混合数，$n = 0, \cdots, N-1$ はそのインデックスである．$h(\omega)$ は周波数ワーピング関数であり，$h(\omega) = \omega$ とすれば $G(\omega; \rho_{r,n}, \nu_{r,n}^2)$ は線形周波数領域で平均 $\rho_{r,n}$，分散 $\nu_{r,n}^2$ の正規分布と同形の関数となる．メル周波数領域で滑らかな声道スペクトルとなるように以下のように $h(\omega)$ を設計した．

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

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．$W = (W_{r,n})_{r,n} \geqq 0$



(a) 調波成分のみを用いた設計例.



(b) 非周期性指標を用いた設計例.

図 2　与えられた F_0 に対する信頼度 $Z_{k,t}$ の設計例.（a）と（b）の上図では濃いほど値が 1 に近い.

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

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

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 定 式 化

STRAIGHT により得ら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 $H^{(\mathrm{GMM})}_{k,r}$ を用いた場合は $\Theta = \{W, U\}$，$H^{(\mathrm{AR})}_{k,r}$ を用いた場合は $\Theta = \{\{\boldsymbol{a}_r\}_r, U\}$ である．以後簡単のため，$H^{(\mathrm{GMM})}_{k,r}$ を用いた場合の NMF を GMM-NMF，$H^{(\mathrm{AR})}_{k,r}$ 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.

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，

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

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

と書ける.

2.3 信頼度の設計

信頼度 $Z_{k,t}$ の単純な設計方法として，各時刻での F_0 とその高調波周波数に対応する時間周波数成分には $Z_{k,t} = 1$，それ以外の時間周波数成分には $Z_{k,t} = 0$ とする方法が考えられる．この方法で作られた $Z_{k,t}$ を図 2 に示す．ただし，STRAIGHT により F_0 が推定されなかった時刻については全ての k で $Z_{k,t} = 1$ とした．音声データが少ない場合にこの方法を用いると，図 2（a）に示した通り実際に学習に使われる時間周波数成分が大幅に少なくなる．そのため，アンダーフィッティングをおこす可能性がある．アンダーフィッティング抑制する方法の 1 つは，$Z_{k,t} = 0$ としたいくつかの時間周波数成分の信頼度を 1 以下の正値にすることである．この値を実験的，経験的に決定することもできるが，本稿では STRAIGHT で得られる非周期性指標 $A_{k,t}$¥$in[0,1]$ を利用して設計する方法を紹介する．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（図 2（b））.

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る．(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 [6], [7] 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．

$$
\begin{aligned}
&-Z_{k,t}Y_{k,t}\ln\left(\sum_{r,n}W_{r,n}G(\omega_k;\rho_{r,n},\nu_{r,n}^2)U_{r,t}\right)\\
&\leqq -Z_{k,t}Y_{k,t}\sum_{r,n}\lambda_{k,t,r,n}\ln\left(\frac{W_{r,n}G(\omega_k;\rho_{r,n},\nu_{r,n}^2)U_{r,t}}{\lambda_{k,t,r,n}}\right)
\end{aligned}
\tag{9}
$$

ここで，$\lambda_{k,t,r,n}\geqq 0$ 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，

$$
\lambda_{k,t,r,n}=\frac{W_{r,n}G(\omega_k;\rho_{r,n},\nu_{r,n}^2)U_{r,t}}{X_{k,t}} \tag{10}
$$

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

$$
\begin{aligned}
&\mathcal{L}^{+}_{\mathrm{GKL,GMM}}(\Theta,\lambda)\\
&=\sum_{k,t}Z_{k,t}Y_{k,t}\left(-\sum_{r,n}\lambda_{k,t,r,n}\ln\frac{W_{r,n}G(\omega_k;\rho_{r,n},\nu_{r,n}^2)U_{r,t}}{\lambda_{k,t,r}}\right.\\
&\quad\left.+\ln Y_{k,t}\right)+\sum_{k,t}Z_{k,t}(X_{k,t}-Y_{k,t})
\end{aligned}
\tag{11}
$$

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．

$$
W_{r,n}\propto W_{r,n}\frac{\sum_{k,t}Z_{k,t}\dfrac{Y_{k,t}}{X_{k,t}}G(\omega_k;\rho_{r,n},\nu_{r,n}^2)U_{r,t}}{\sum_{k,t}Z_{k,t}G(\omega_k;\rho_{r,n},\nu_{r,n}^2)U_{r,t}} \tag{12}
$$

$$
U_{r,t}\leftarrow U_{r,t}\frac{\sum_{k}Z_{k,t}\dfrac{Y_{k,t}}{X_{k,t}}H_{k,r}^{(\mathrm{GMM})}}{\sum_{k}Z_{k,t}H_{k,r}^{(\mathrm{GMM})}} \tag{13}
$$

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．

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，

$$
\begin{aligned}
&\mathcal{L}^{+}_{\mathrm{EU,GMM}}(\Theta,\{\eta_{k,t,r,n}\}_{k,t,r,n})\\
&=\sum_{k,t}Z_{k,t}(Y_{k,t}^2-2Y_{k,t}X_{k,t})\\
&\quad+\sum_{k,t,r,n}\frac{Z_{k,t}(W_{r,n}G(\omega_k;\rho_{r,n},\nu_{r,n}^2)U_{r,t})^2}{\eta_{k,t,r,n}}
\end{aligned}
\tag{14}
$$

と補助関数を設計できる．ここで，$\eta_{k,t,r,n}\geqq 0$ は補助変数であり，各 k,t に関し $\sum_{r,n}\eta_{k,t,r,n}=1$ を満たす．等号成立条件は

$$
\eta_{k,t,r,n}=\frac{W_{r,n}G(\omega_k;\rho_{r,n},\nu_{r,n}^2)U_{r,t}}{X_{k,t}} \tag{15}
$$

である．上述の場合と同様に更新式は以下のように得られる．

$$
W_{r,n}\leftarrow W_{r,n}\frac{\sum_{k,t}Z_{k,t}Y_{k,t}G(\omega_k;\rho_{r,n},\nu_{r,n}^2)U_{r,t}}{\sum_{k,t}Z_{k,t}X_{k,t}G(\omega_k;\rho_{r,n},\nu_{r,n}^2)U_{r,t}} \tag{16}
$$

$$
U_{r,t}\leftarrow U_{r,t}\frac{\sum_{k}Z_{k,t}Y_{k,t}H_{k,r}^{(\mathrm{GMM})}}{\sum_{k}Z_{k,t}X_{k,t}H_{k,r}^{(\mathrm{GMM})}} \tag{17}
$$

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，

$$
\begin{aligned}
&\mathcal{L}^{+}_{\mathrm{GKL,AR}}(\Theta,\xi)\\
&=\sum_{k,t}Z_{k,t}Y_{k,t}\left(\ln Y_{k,t}-\sum_{r}\xi_{k,t,r}\ln\frac{H_{k,r}^{(\mathrm{AR})}U_{r,t}}{\xi_{k,t,r}}\right)\\
&\quad+\sum_{k,t}Z_{k,t}(X_{k,t}-Y_{k,t})
\end{aligned}
\tag{18}
$$

と定義できる．等式成立条件は $\xi_{k,t,r}=H_{k,r}^{(\mathrm{AR})}U_{r,t}/X_{k,t}$ である．$U_{r,t}$ の更新式は，(13) 式の $H_{k,r}^{(\mathrm{GMM})}$ を $H_{k,r}^{(\mathrm{AR})}$ 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 [8], [9]. $\mathcal{L}_{\mathrm{GKL}}(\Theta)$ の $\boldsymbol{a}_r$ に関する偏微分は，

$$
\frac{\partial\mathcal{L}_{\mathrm{GKL}}}{\partial\boldsymbol{a}_r}(\Theta)=(\Psi_r^{+}-\Psi_r^{-})\boldsymbol{a}_r \tag{19}
$$

$$
\Psi_r^{+}=\sum_{k,t}Z_{k,t}\frac{Y_{k,t}}{X_{k,t}}(H_{k,r}^{(\mathrm{AR})})^3U_{r,t}Q(\omega_k) \tag{20}
$$

$$
\Psi_r^{-}=\sum_{k,t}Z_{k,t}(H_{k,r}^{(\mathrm{AR})})^3U_{r,t} \tag{21}
$$

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．

$$
\boldsymbol{a}_r\leftarrow(\Psi_r^{+})^{-1}\Psi_r^{-}\boldsymbol{a}_r. \tag{22}
$$

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，(17) 式の $H_{k,r}^{(\mathrm{GMM})}$ を $H_{k,r}^{(\mathrm{AR})}$ に置換したものと同一であり，$\boldsymbol{a}_r$ の更新式は (22) 式と同様に以下で与えられる．

$$
\begin{aligned}
\boldsymbol{a}_r\leftarrow&\left(\sum_{k,t}Z_{k,t}Y_{k,t}(H_{k,r}^{(\mathrm{AR})})^3U_{r,t}Q(\omega_k)\right)^{-1}\\
&\times\left(\sum_{k,t}Z_{k,t}X_{k,t}(H_{k,r}^{(\mathrm{AR})})^3U_{r,t}Q(\omega_k)\right)\boldsymbol{a}_r
\end{aligned}
\tag{23}
$$

4. 声道スペクトル推定精度評価実験

4.1 実験条件

[4] 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 [10] 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

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

表 2　F_0 を x 倍された男性話者の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]．声道スペクトル推定法は表 1 と同一であり，括弧内の値は非周期性指標を用いない場合の結果である．

x	GKL	EU	STRAIGHT
$2^{-1.0}$	$\mathbf{4.53 \pm 1.82}$	5.56 ± 1.71	4.63 ± 1.99
$2^{-0.5}$	$\mathbf{3.91 \pm 1.91}$	5.05 ± 1.77	3.98 ± 2.16
$2^{0.0}$	$\mathbf{3.78 \pm 1.89}$	4.97 ± 1.91	3.82 ± 2.17
$2^{0.5}$	$\mathbf{4.04 \pm 1.65}$	5.09 ± 1.79	4.21 ± 1.88
$2^{1.0}$	$\mathbf{4.71 \pm 1.47}$	5.39 ± 1.59	4.89 ± 1.59
$2^{1.5}$	$\mathbf{5.64 \pm 1.54}$	6.33 ± 1.92	5.80 ± 1.58

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 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，$r = 0, \cdots, R, n = 0, \cdots, N-1$ に対して $\rho_{r,n} = n/(N-1)$,$\nu_{r,n} = 1/(N-1)$ 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，日本人男性話者 1 名による 50 文の音声信号に関しても同様に実験を行った．女性話者と比べ男性話者は F_0 が低いため，女性話者の場合の結果に比べ観測可能な調波成分が多い．そのため，女性話者の場合に比べ STRAIGHT からの性能の向上が小さいことが予想される．実際に表 2 に示す通り，提案法を用いることで平均的にはメルケプストラム歪みは少なくなったが，女性話者での結果に比べ性能の向上は小さかった．F_0 以外にも，男性話者の音声信号のフレーム数が女性話者の半分程度だったことも，メルケプストラム歪みでの改善量が少なかったことの原因の 1 つであると考えられる．今後は，メルケプストラム歪みに対するフレーム数の効果を検証するために様々なデータ数での評価実験が必要である．

5. 結　　論

本稿では，声道スペクトログラムの大局的な手がかりを利用し，各フレームでの調波成分を重視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が低ランクな非負値行列で近似できると仮定し，声道スペクトル推定問題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 に比べ平均的に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，フレーム数を増やしての大規模実験や性能評価，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．

文　　献

[1] H. Kawahara, I. Masuda-Katsuse, and A. deCheveigné, “Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds,” Speech Commun., vol.27, no.3, pp.187–207, 1999.

[2] H. Kawahara, M. Morise, T. Takahashi, R. Nisimura, T. Irino, and H. Banno, “TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation,” Proc. Int. Conf. Acoust. Speech Signal Process., pp.3933–3936, 2008.

[3] Y. Shiga and S. King, “Estimating the spectral envelope

of voiced speech using multi-frame analysis," Proc. EUROSPEECH, pp.1737–1740, 2003.

[4] T. Toda, "Statistical approach to vocal tract transfer function estimation based on factor analyzed trajectory hmm," Proc. Int. Conf. Acoust. Speech Signal Process., pp.3925–3928, 2008.

[5] H. Kameoka, N. Ono, and S. Sagayama, "Speech spectrum modeling for joint estimation of spectral envelope and fundamental frequency," IEEE Trans. Acoust., Speech, and Language Process., vol.18, no.6, pp.1507–1516, 2010.

[6] J.M. Ortega and W.C. Rheinboldt, Iterative solution of nonlinear equations in several variables, SIAM, 1970.

[7] D.R. Hunter and K. Lange, "Quantile regression via an MM algorithm," J. Comp. Graph. Stat., vol.9, no.1, pp.60–77, 2000.

[8] A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling," IEEE Trans. Signal Process., vol.39, no.2, pp.411–423, Feb. 1991.

[9] R. Badeau, N. Bertin, and E. Vincent, "Stability analysis of multiplicative update algorithms and application to nonnegative matrix factorization," IEEE Trans. Neural Netw., vol.21, no.12, pp.1869–1881, 2010.

[10] A. Kurematsu, K. Takeda, Y. Sagisaka, S. Katagiri, H. Kuwabara, and K. Shikano, "ATR Japanese speech database as a tool of speech recognition and synthesis," Speech Commun., vol.9, no.4, pp.357–363, 1990.