トラジェクトリ隠れマルコフモデルによる音声強調 *

☆岸田拓也 (九州大), 亀岡弘和 (NTT/国立情報学研), 中島祥好 (九州大)

1　はじめに

音声信号に混入する雑音を抑圧する音声強調技術は，高品質な音声通信を達成するためや音声認識の精度を上げるための重要技術の一つである．どのような雑音が混入するのかを予測できない未知雑音環境下においても，例えば特定話者の利用を想定した音声通信システムのように，強調すべき音声の学習サンプルを事前に用意することが可能な場面は多い．このような状況での音声強調手法として，半教師あり非負値行列因子分解 (Semi-supervised Non-negative Matrix Factorization: SSNMF) に基づく手法が提案されている [1]．この手法は，各時刻の観測スペクトルを事前学習した音声の基底スペクトルを並べた行列と雑音の基底スペクトルを並べた行列の非負結合でフィッティングすることで音声と雑音のパワースペクトルを推定するというものである．SSNMF 法は未知雑音環境下においても高い信号対雑音比の強調音声を得られる一方で，残留雑音成分からなるミュージカルノイズ（時間周波数平面に点在する雑音成分）が原因で必ずしも聴感的に品質の良い音声とならない場合が多い．また，この SSNMF 法は音声のスペクトルの時間変化の自然さを保証しないため，強調音声のスペクトルが不連続になりがちであり，これもまた強調音声の聴感上の品質を下げる原因となっている．

以上のような問題の解消のために，音声の聴こえをより自然にするためのアプローチを音声強調の手法に組み込む必要があると考え，本研究では音声合成の分野で用いられているある手法に着目した．音声合成においては，聴感的に高品質な音声を合成することを目的として対数スペクトルだけでなく対数スペクトルの時間微分量の統計分布を用いて最適な音声スペクトル系列を生成する方式，隠れマルコフモデル (Hidden Markov Model; HMM) 音声合成という手法がとられる [2, 3]．音声通信のための音声強調においても，いかに聴感的に高品質な音声にできるかが重要となるため，音声合成に用いられているモデルや手法は音声強調にも有用であると考えられる．

そこで本研究では，SSNMF 法における音声スペクトル系列のモデルとして，従来の行列積型のモデルの代わりに，HMM 音声合成の生成モデルと同形のモデルを用いた音声強調手法を提案する．

2　提案モデルの定式化

観測信号の振幅スペクトルまたはパワースペクトル（以後，観測スペクトル）を $Y_{\omega,t}$ とする．ただし，ω と t はそれぞれ周波数と時刻のインデックスである．スペクトルの加法性を仮定し，各時刻の音声スペクトル $X^{(S)}_{\omega,t}$ および雑音スペクトル $X^{(N)}_{\omega,t}$ をそれぞれ L_S 個の基底スペクトル $H^{(S)}_{\omega,1},\ldots,H^{(S)}_{\omega,L_S}$ と L_N 個の基底スペクトル $H^{(N)}_{\omega,1},\ldots,H^{(N)}_{\omega,L_N}$ の非負結合

$$X^{(S)}_{\omega,t}=\sum_{l=1}^{L_S}H^{(S)}_{\omega,l}U^{(\mathrm{S})}_{l,t} \tag{1}$$

$$X^{(N)}_{\omega,t}=\sum_{l=1}^{L_N}H^{(N)}_{\omega,l}U^{(N)}_{l,t} \tag{2}$$

で表せるものとする．SSNMF 法は，クリーン音声の学習サンプルから事前学習した $H^{(S)}_{\omega,1},\ldots,H^{(S)}_{\omega,L_S}$ を用いて，観測スペクトル $Y_{\omega,t}$ に $X_{\omega,t}=X^{(S)}_{\omega,t}+X^{(N)}_{\omega,t}$ をフィッティングすることで観測スペクトログラムに含まれる音声の成分と雑音の成分を推定する方法である [1]．$Y_{\omega,t}$ に対する $X_{\omega,t}$ のフィッティングの度合いは，例えば，

$$\mathcal{I}(Y|X)=\sum_{\omega,t}(Y_{\omega,t}-X_{\omega,t})^2 \tag{3}$$

のような規準で測ることができる．しかしながら，音声の基底スペクトルで雑音スペクトルを説明できてしまう場合やその逆の場合がありえるため，$Y_{\omega,t}$ と $X_{\omega,t}$ の誤差を小さくできたとしても $X^{(S)}_{\omega,t}$ と $X^{(N)}_{\omega,t}$ が実際の音声スペクトルと雑音スペクトルに対応するとは限らない．また，この方式では，音声スペクトルの推定において音声スペクトルの時間変化量の統計を活用できる仕組みになっておらず，このことがミュージカルノイズの発生や不連続なスペクトル変化を許す一因になっていたと考えられる．そこでより高品質な音声強調を実現するためには音声スペクトルの時間変化の傾向を考慮し，同じ $X_{\omega,t}$ を与える $X^{(S)}_{\omega,t}$ と $X^{(N)}_{\omega,t}$ の不定性を解消するより強い制約が必要である．

そこで，本稿では，$\mathcal{J}(Y|X)$, $\mathcal{K}_F(F|\boldsymbol{s})$, $\mathcal{K}_E(E|\boldsymbol{q})$, $\mathcal{L}(X^{(N)}|H^{(N)},U^{(N)})$ の 4 つの規準の重みつき和

$$\begin{aligned}\mathcal{I}(F,E,H^{(N)},U^{(N)})&=\alpha_1\mathcal{J}(Y|X)\\&\quad+\alpha_2\{\mathcal{K}_F(F|\boldsymbol{s})+\mathcal{K}_E(E|\boldsymbol{q})\}\\&\quad+\alpha_3\mathcal{L}(X^{(N)}|H^{(N)},U^{(N)})\end{aligned} \tag{4}$$

を目的関数とした最適化問題を定式化し，停留点への収束が保証されるパラメータ推定アルゴリズムを提案する．ただし，α_1, α_2, α_3 は重みづけ係数である．各項について以下で述べる．

式 (4) の右辺第一項の $\mathcal{J}(Y|X)$ は観測スペクトル Y と推定した音声と雑音のスペクトルの和 X の誤差であり，以下の対数スペクトル距離規準を用いる．

$$\mathcal{J}(Y|X)=\sum_{\omega,t}\left(\log Y_{\omega,t}-\log X_{\omega,t}\right)^2 \tag{5}$$

*Speech Enhancement Based on Trajectory Hidden Markov Model. by KISHIDA, Takuya (Kyushu University), KAMEOKA, Hirokazu (NTT/National Institute of Informatics), and NAKAJIMA, Yoshitaka (Kyushu University)

第二項は，音声のスペクトルを HMM 音声合成の生成モデルと同形のモデルによって説明するための規準となる項であり，

$$\mathcal{K}_F(F|\boldsymbol{s}) = \sum_{\omega,t}\big(\log F_{\omega,t} - \rho_\omega(\boldsymbol{s})\big)^2 \tag{6}$$

$$\mathcal{K}_E(E|\boldsymbol{q}) = \sum_{\omega,t}\big(\log E_{\omega,t} - \mu_\omega(\boldsymbol{q})\big)^2 \tag{7}$$

である．ここで，音声信号のスペクトル $X^{(S)}_{\omega,t}$ を，音声のソースフィルタ理論に基づき，$X^{(S)}_{\omega,t} = F_{\omega,t}E_{\omega,t}$ と声道スペクトル $F_{\omega,t}$ と声帯振動スペクトル $E_{\omega,t}$ に分解している．また，$\boldsymbol{s} = (s_1,\dots,s_T)$ および $\boldsymbol{q} = (q_1,\dots,q_T)$ はそれぞれ声道スペクトル系列と声帯振動スペクトル系列に対応した異なる HMM の状態系列を表し，$\boldsymbol{\rho}_\omega(\boldsymbol{s}) = (\rho_{\omega,1}(\boldsymbol{s}),\dots,\rho_{\omega,T}(\boldsymbol{s}))^\mathsf{T}$ および $\boldsymbol{\mu}_\omega(\boldsymbol{q}) = (\mu_{\omega,1}(\boldsymbol{q}),\dots,\mu_{\omega,T}(\boldsymbol{q}))^\mathsf{T}$ は状態系列が与えられた下での最尤スペクトル系列

$$\boldsymbol{\rho}_\omega(\boldsymbol{s}) = \underset{\boldsymbol{z}_\omega}{\mathrm{argmax}}\,\mathcal{N}(\boldsymbol{W}\boldsymbol{z}_\omega;\boldsymbol{m}_\omega(\boldsymbol{s}),\boldsymbol{V}_\omega(\boldsymbol{s})) \tag{8}$$

$$\boldsymbol{\mu}_\omega(\boldsymbol{q}) = \underset{\boldsymbol{z}_\omega}{\mathrm{argmax}}\,\mathcal{N}(\boldsymbol{W}\boldsymbol{z}_\omega;\boldsymbol{c}_\omega(\boldsymbol{q}),\boldsymbol{U}_\omega(\boldsymbol{q})) \tag{9}$$

に対応する．これは HMM 音声合成におけるパラメータ生成方式 [2, 3] と同形で，

$$\boldsymbol{\rho}_\omega(\boldsymbol{s}) = (\boldsymbol{W}^\mathsf{T}\boldsymbol{V}_\omega(\boldsymbol{s})\boldsymbol{W})^{-1}\boldsymbol{W}^\mathsf{T}\boldsymbol{V}_\omega(\boldsymbol{s})^{-1}\boldsymbol{m}_\omega(\boldsymbol{s}) \tag{10}$$

$$\boldsymbol{\mu}_\omega(\boldsymbol{q}) = (\boldsymbol{W}^\mathsf{T}\boldsymbol{U}_\omega(\boldsymbol{q})\boldsymbol{W})^{-1}\boldsymbol{W}^\mathsf{T}\boldsymbol{U}_\omega(\boldsymbol{q})^{-1}\boldsymbol{c}_\omega(\boldsymbol{q}) \tag{11}$$

で与えられる．ただし，$\boldsymbol{m_s}$ および $\boldsymbol{c_q}$ は状態系列 $\boldsymbol{s}$, $\boldsymbol{q}$ が与えられた下での各々の HMM の状態出力分布の平均系列，$\boldsymbol{V_s}$ および $\boldsymbol{U_q}$ は状態系列 $\boldsymbol{s}$, $\boldsymbol{q}$ が与えられた下での各々の HMM の状態出力分布の分散を対角成分に並べた対角行列である．また，$\boldsymbol{W}$ は，パラメータの時系列が格納されたベクトルをそれ自身とその時間数値微分系列を結合したベクトルに変換する行列である．また，$\mathcal{N}(x;\mu,\Sigma)$ は平均が μ，分散共分散行列が Σ の正規分布を表す．従って式 (8), (9) は所与の状態系列と状態出力分布によって決まるスペクトル系列とその時間微分系列の両確率分布を規準とした最尤スペクトル系列を意味していることが分かる．

第三項は，従来の SSNMF 法と同様に，雑音スペクトログラム $X^{(N)}_{\omega,t}$ を行列積型のモデルで説明するための規準であり，以下のようにした．

$$\begin{aligned}&\mathcal{L}(X^{(N)}|H^{(N)},U^{(N)})\\ &= \sum_{\omega,t}\Big(\log X^{(N)}_{\omega,t} - \log\sum_l H^{(N)}_{\omega,l}U^{(N)}_{l,t}\Big)^2\end{aligned} \tag{12}$$

3 パラメータ推定アルゴリズム

$\mathcal{J}(F,E,H^{(N)},U^{(N)})$ を最小化する F, E, $H^{(N)}$,$U^{(N)}$ を解析的に得ることはできないが，当該最適化問題の局所最適解を探索する反復アルゴリズムを補助関数法に基づき導くことができる．

補助関数法による，目的関数 $F(\theta)$ の最小化問題の最適化アルゴリズムでは，まず補助変数 ξ を導入し，$F(\theta) = \min_\xi F^+(\theta,\xi)$ を満たす補助関数 $F^+(\theta,\xi)$ を設計する．このような補助関数が設計できれば，$\xi \leftarrow \mathrm{argmin}_\xi F^+(\theta,\xi)$ と $\theta \leftarrow \mathrm{argmin}_\theta F^+(\theta,\xi)$ を交互に繰り返すことで，目的関数 $F(\theta)$ を局所最小化する θ を得ることができる．本最適化問題における目的関数の補助関数は，目的関数の中の凸関数と凹関数に対してそれぞれ Jensen の不等式と接線不等式を利用し，さらに対数の二乗の関数については，$x > 0, \xi > 0$ において成り立つ以下の不等式

$$(\log x)^2 \le \frac{1}{x} + f(\xi)x + g(\xi) \tag{13}$$

$$f(\xi) = \frac{2\log\xi}{\xi} + \frac{1}{\xi^2} \tag{14}$$

$$g(\xi) = (\log\xi)^2 - 2\log\xi - \frac{2}{\xi} \tag{15}$$

を利用して設計することができる [4]．まず，$\mathcal{J}(Y|X)$ については，

$$\begin{aligned}&\mathcal{J}(Y|X)\\ &= \sum_{\omega,t}(\log X_{\omega,t})^2\\ &- 2\sum_{\omega,t}\delta_{\log Y\ge 0}|\log Y_{\omega,t}|\log X_{\omega,t}\\ &+ 2\sum_{\omega,t}\delta_{\log Y<0}|\log Y_{\omega,t}|\log X_{\omega,t}\\ &+ \sum_{\omega,t}(\log Y_{\omega,t})^2\end{aligned} \tag{16}$$

$$\begin{aligned}&\le \sum_{\omega,t}\left\{\frac{(\lambda^{(S)}_{\omega,t})^2}{X^{(S)}_{\omega,t}} + \frac{(\lambda^{(N)}_{\omega,t})^2}{X^{(N)}_{\omega,t}} + f(\xi_{\omega,t})X_{\omega,t} + g(\xi_{\omega,t})\right\}\\ &- 2\sum_{\omega,t}\delta_{\log Y\ge 0}|\log Y_{\omega,t}|\left(\theta^{(S)}_{\omega,t}\log\frac{X^{(S)}_{\omega,t}}{\theta^{(S)}_{\omega,t}} + \theta^{(N)}_{\omega,t}\log\frac{X^{(N)}_{\omega,t}}{\theta^{(N)}_{\omega,t}}\right)\\ &+ 2\sum_{\omega,t}\delta_{\log Y<0}|\log Y_{\omega,t}|\left(\frac{X_{\omega,t}}{\phi_{\omega,t}} + \log\phi_{\omega,t} - 1\right)\\ &+ \sum_{\omega,t}(\log Y_{\omega,t})^2\end{aligned} \tag{17}$$

のように補助関数を作ることができる．ただし，δ_x は条件 x を満たす場合に 1，満たさない場合に 0 となる指示関数である．また，式 (17) の等号が成立するのは，補助変数が

$$\lambda^{(S)}_{\omega,t} = \frac{X^{(S)}_{\omega,t}}{X_{\omega,t}},\quad \lambda^{(N)}_{\omega,t} = \frac{X^{(N)}_{\omega,t}}{X_{\omega,t}} \tag{18}$$

$$\xi_{\omega,t} = X_{\omega,t} \tag{19}$$

$$\theta^{(S)}_{\omega,t} = \frac{X^{(S)}_{\omega,t}}{X_{\omega,t}},\quad \theta^{(N)}_{\omega,t} = \frac{X^{(N)}_{\omega,t}}{X_{\omega,t}} \tag{20}$$

$$\phi_{\omega,t} = X_{\omega,t} \tag{21}$$

を満たすときである．

次に，$\mathcal{K}_F(F|\boldsymbol{s})$ と $\mathcal{K}_E(E|\boldsymbol{q})$ については，$\eta_{\omega,t} > 0$,

$\zeta_{\omega,t}>0$ を用いて

$$\mathcal{K}_F(F|\boldsymbol{s})=\sum_{\omega,t}\big(\log F_{\omega,t}-\rho_\omega(\boldsymbol{s})\big)^2$$
$$=\sum_{\omega,t}\Big\{(\log F_{\omega,t})^2-2\log F_{\omega,t}\rho_\omega(\boldsymbol{s})+\{\rho_\omega(\boldsymbol{s})\}^2\Big\}$$
$$\leq\sum_{\omega,t}\left\{\frac{1}{F_{\omega,t}}+f(\eta_{\omega,t})F_{\omega,t}+g(\eta_{\omega,t})\right\}$$
$$-2\sum_{\omega,t}\log F_{\omega,t}\rho_\omega(\boldsymbol{s})+\sum_{\omega,t}\{\rho_\omega(\boldsymbol{s})\}^2 \tag{22}$$

$$\mathcal{K}_E(E|\boldsymbol{q})=\sum_{\omega,t}\big(\log E_{\omega,t}-\mu_\omega(\boldsymbol{q})\big)^2$$
$$=\sum_{\omega,t}\Big\{(\log E_{\omega,t})^2-2\log E_{\omega,t}\mu_\omega(\boldsymbol{q})+\{\mu_\omega(\boldsymbol{q})\}^2\Big\}$$
$$\leq\sum_{\omega,t}\left\{\frac{1}{E_{\omega,t}}+f(\zeta_{\omega,t})E_{\omega,t}+g(\zeta_{\omega,t})\right\}$$
$$-2\sum_{\omega,t}\log E_{\omega,t}\mu_\omega(\boldsymbol{q})+\sum_{\omega,t}\{\mu_\omega(\boldsymbol{q})\}^2 \tag{23}$$

のような不等式が立てられ，それぞれの右辺を補助関数とすることができる．なお，等号は

$$\eta_{\omega,t}=F_{\omega,t},\quad \zeta_{\omega,t}=E_{\omega,t} \tag{24}$$

のとき成り立つ．

最後に，$\mathcal{L}(X^{(N)}|H^{(N)},U^{(N)})$ については，

$$\mathcal{L}(X^{(N)}|H^{(N)},U^{(N)})$$
$$=\sum_{\omega,t}\Big(\log X^{(N)}_{\omega,t}-\log\sum_l H^{(N)}_{\omega,l}U^{(N)}_{l,t}\Big)^2$$
$$=\sum_{\omega,t}(\log X^{(N)}_{\omega,t})^2$$
$$-2\sum_{\omega,t}\log X^{(N)}_{\omega,t}\left(\log\sum_l H^{(N)}_{\omega,l}U^{(N)}_{l,t}\right)$$
$$+\sum_{\omega,t}\left(\log\sum_l H^{(N)}_{\omega,l}U^{(N)}_{l,t}\right)^2$$
$$\leq\sum_{\omega,t}\left\{\frac{1}{X^{(N)}_{\omega,t}}+f(\nu_{\omega,t})X^{(N)}_{\omega,t}+g(\nu_{\omega,t})\right\}$$
$$-2\sum_{\omega,t}\delta_{\log X^{(N)}\geq 0}|\log X^{(N)}_{\omega,t}|\sum_l\psi_{\omega,l,t}\log\frac{H^{(N)}_{\omega,l}U^{(N)}_{l,t}}{\psi_{\omega,l,t}}$$
$$+2\sum_{\omega,t}\delta_{\log X^{(N)}<0}|\log X^{(N)}_{\omega,t}|\left(\frac{\sum_l H^{(N)}_{\omega,l}U^{(N)}_{l,t}}{\pi_{\omega,t}}+\log\pi_{\omega,t}-1\right)$$
$$+\sum_{\omega,t}\left\{\sum_l\frac{\tau^2_{\omega,l,t}}{H^{(N)}_{\omega,l}U^{(N)}_{l,t}}+f(\sigma_{\omega,t})\sum_l H^{(N)}_{\omega,l}U^{(N)}_{l,t}+g(\sigma_{\omega,t})\right\} \tag{25}$$

のような不等式が立てられ，右辺を補助関数とすることができる．なお，式 (25) の等号成立条件は以下となる．

$$\nu_{\omega,t}=X^{(N)}_{\omega,t} \tag{26}$$

$$\psi_{\omega,l,t}=\tau_{\omega,l,t}=\frac{H^{(N)}_{\omega,l}U^{(N)}_{l,t}}{\sum_l H^{(N)}_{\omega,l}U^{(N)}_{l,t}} \tag{27}$$

$$\pi_{\omega,t}=\sigma_{\omega,t}=\sum_l H^{(N)}_{\omega,l}U^{(N)}_{l,t} \tag{28}$$

以上の補助関数を足し合わせることで $\mathcal{J}(F,E,H^{(N)},U^{(N)})$ の補助関数が得られる．この補助関数の各パラメータに関する偏微分が 0 となるパラメータは二次方程式を解くことで得られる．$\rho_{\omega,t}(\boldsymbol{s})$ の更新については，まず，$F_{\omega,1},\dots,F_{\omega,T}$ の更新後，$F_{\omega,1},\dots,F_{\omega,T}$ の時間微分系列 $\Delta F_{\omega,1},\dots,\Delta F_{\omega,T}$ を算出し，$F_{\omega,t}$ と $\Delta F_{\omega,t}$ 結合ベクトルを $\boldsymbol{o}_{\omega,t}=(F_{\omega,t},\Delta F_{\omega,t})^{\mathsf{T}}$ とする．$\boldsymbol{o}_{\omega,1}\dots,\boldsymbol{o}_{\omega,T}$ を HMM の出力系列と見なし，その生成確率が最大となる状態系列 $\boldsymbol{s}$ を Viterbi アルゴリズムにより探索する．求まった $\boldsymbol{s}$ を用いて式 (8) により新しい $\rho_{\omega,t}(\boldsymbol{s})$ を得る．以上と同様の手続きを $E_{\omega,1},\dots,E_{\omega,T}$ にも適用し，$\boldsymbol{q}$ と $\mu_{\omega,t}(\boldsymbol{q})$ を更新する．

4 性能評価実験

4.1 実験条件

NTT-AT 多言語音声データベース 2002 の音声データと RWCP の雑音データ (museum noise, babble noise, background music noise の 3 種類) および白色雑音を用いて上述の手法による雑音抑圧効果を検証する評価実験を行った．比較対象は従来の SSNMF 法とし，強調前と強調後の信号対雑音比（SNR）を評価した．テストデータは 6 名 (うち女性 3 名，男性 3 名) の話者のクリーン音声に各雑音を-10, -5, 0 dB の三種類の SNR で重畳させて作成した．テストデータはすべてサンプリング周波数 16 kHz のモノラル信号で，フレーム長 32 ms, フレームシフト 16 ms で短時間 Fourier 変換を行い，観測スペクトログラム $Y_{\omega,t}$ を算出した．学習においてはテストデータに用いる 6 名を含む 10 名 (うち女性 5 名，男性 5 名) の話者の計 500 文の音声を用いて $H^{(S)}_{\omega,l}$ を学習し，音声の声道スペクトルと声帯振動スペクトルに対して，その対数スペクトル系列および対数スペクトルの一次時間数値微分系列の HMM パラメータの学習を行った．なお，モデルの簡略化のため，HMM の状態遷移確率は考えないこととした．事前学習した $H^{(S)}_{\omega,l}$ の基底数は 6 とし，HMM の状態数は 32 とした．テストにおいては，学習で得られた $H^{(S)}_{\omega,l}$ と HMM パラメータを固定し，$\alpha_1=10,\alpha_2=1,\alpha_3=10$ として $F_{\omega,t},E_{\omega,t},X^{(N)}H^{(N)}_{\omega,l},U^{(N)}_{l,t},\boldsymbol{s},\boldsymbol{q},\rho_\omega,\mu_\omega$ の推定を行った．パラメータ推定アルゴリズムの更新回数は 500 回としたが，$H^{(N)}_{\omega,l},U^{(N)}_{l,t},\boldsymbol{s},\boldsymbol{q},\rho_\omega,\mu_\omega$ は 50 回に 1 回だけ更新した．推定後，$X^{(S)}_{\omega,t}$ と $X^{(N)}_{\omega,t}$ を用いて Wiener フィルタにより音声信号の推定値を算出した．アルゴリズムの初期値は従来の SSNMF により得た．

4.2 結果と考察

強調前と強調後の音声のスペクトログラムの例を Fig. 1 に示す．強調後のスペクトログラムは提案法の方が元の音声のスペクトログラムにより近づいていることが確認できる．Fig 2 は強調後の SNR の改善値を示したものである．SNR が-5, 0 dB の条件では，全ての雑音で提案法の方が従来法よりも SNR の改善値の平均値が大きかった．一方で，SNR が-10 dB の条件では，白色雑音以外の雑音の場合，提案法による SNR の改善値の平均値は従来法に下回った．

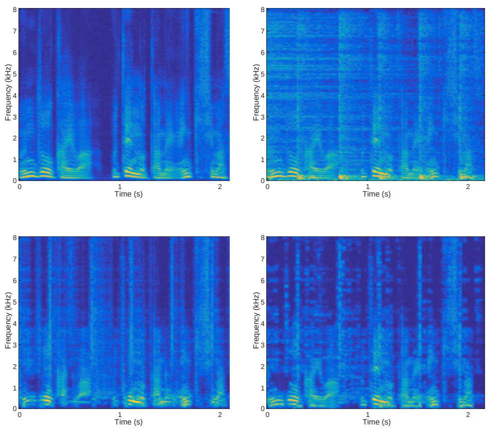

Fig. 1　強調前，強調後の信号のスペクトログラム．原音声 (左上)，原音声に背景音楽雑音を重畳したもの (右上)，従来法による強調音声 (左下)，そして提案法による強調後音声 (右下)．

本性能評価実験では，アルゴリズムの実装においてHMM音声合成モデルの状態遷移確率を考えない簡略化されたモデルを用いたものの，SNRが0 dBである場合には，提案法による音声強調が従来法よりも圧倒的に優れていることが分かる結果が得られている．よって本提案モデルによる音声強調が有効に働く可能性が十分にあることを本性能評価実験で示すことができたと言える．

謝辞　本研究の一部は，JSPS 科研費 26730100 および 26280060，学術振興会特別研究員奨励費 16J05172 の支援を受けた．

参考文献

[1] P. Smaragdis, B. Raj, and M. Shashanka, “Supervised and semi-supervised separation of sounds from single-channel mixtures,” in Proc. Independent Component Analysis and Signal Separation, pp. 414–421, 2007.

[2] T. Yoshimura, K. Tokuda, T. Masuko, T. Kobayashi, T. Kitamura, “Simultaneous modeling of spectrum, pitch and duration in HMM-based speech synthesis,” in Proc. European Conference on Speech Communication and Technology (EUSIPCO 1999), vol. 5. pp. 2347–2350, 1999.

[3] H. Zen, K. Tokuda, T. Kitamura, “Reformulating the HMM as a trajectory model by imposing explicit relationships between static and dynamic feature vector sequences,” Computer Speech and Language, vol. 21, pp. 153–173, 2007.

[4] H. Kameoka, M. Nakano, K. Ochiai, Y. Imoto, K. Kashino, S. Sagayama, “Constrained and regularized variants of non-negative matrix factorization incorporating music-specific constraints,” in Proc. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2012), pp. 5365–5368, 2012.

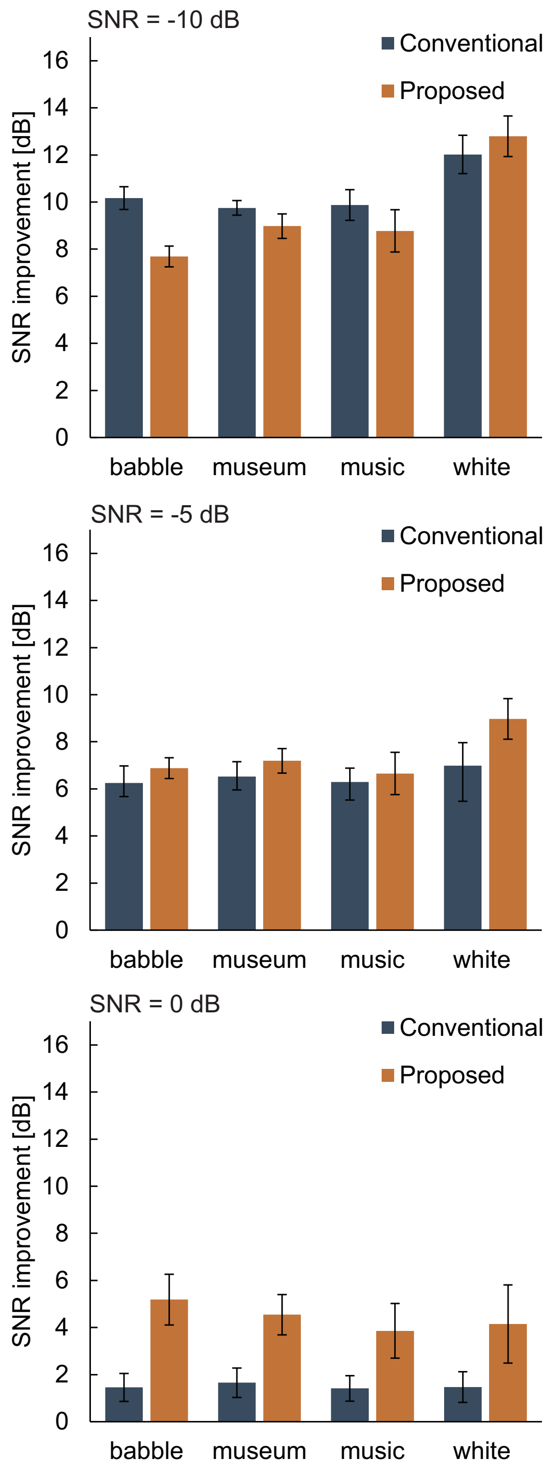

Fig. 2　SNR の改善値．4 種類の雑音信号 (バブル雑音，実環境雑音，背景音楽雑音，白色雑音) を異なる SNR(-10, -5, 0 dB) で音声信号に重畳した信号を SSNMF 法と提案法で強調した．エラーバーは標準偏差．