

ケプストラム距離正則化半教師あり NMF による音声強調*

☆李莉¹, 亀岡弘和^{1,2}, 樋口卓哉², 猿渡洋¹

¹ 東京大学 大学院情報理工学系研究科

² 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

1 はじめに

本稿は音声信号から雑音を抑圧する問題を扱う。音声信号に混入する雑音は音声通信の品質を劣化させるだけでなく音声認識や音声変換などのさまざまな音声処理の性能低下を招く。この問題を解決するためこれまでさまざまな音声強調手法が提案されてきた [1]。

音声強調手法は教師なしアプローチ、教師ありアプローチ、半教師ありアプローチに大別される。教師ありアプローチは対象音声と対象雑音のサンプルが事前に得られる状況、半教師ありアプローチは対象音声のサンプルのみが事前に得られる状況、教師なしアプローチはいずれも得られない状況をそれぞれ想定した音声強調手法である。また、強調する対象が信号（またはスペクトル）の場合と特徴量の場合とでも大別される。教師あり特徴量強調アプローチの代表例としては Vector Taylor Series (VTS) 法 [2, 3], Stereo Piecewise Linear Compensation for Environment (SPLICE) [4, 5], Denoising Autoencoder (DAE) を用いた手法 [6, 7] などがある。VTS 法は、音声と雑音の線形な重畳過程を特徴量空間で 1 次近似することにより雑音あり音声特徴量からクリーン音声特徴量への変換関数を構成する手法である。SPLICE は、雑音あり音声とクリーン音声の特徴量の同時確率密度関数を混合正規分布 (Gaussian Mixture Model: GMM) でモデル化し、学習サンプルを用いて学習した GMM パラメータにより雑音あり音声特徴量からクリーン音声特徴量への変換関数を構成する手法である。DAE 法は、雑音あり音声特徴量を入力、クリーン音声特徴量を出力とした深層ニューラルネットワークにより入出力間の変換関数を構成する手法である。これら教師あり音声強調アプローチは、識別モデルや識別的規準に基づくため、既知の雑音環境下では極めて強力であるが、未知の雑音環境下では必ずしも有効ではない。ただし、学習データの音声または雑音がテスト時のものと異なる場合にそのミスマッチを補償する方法も多く提案されている。

一方、半教師あり信号強調アプローチの代表例である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 [8] は、未知の雑音環境下における強力な音声強調法として近年注目されている。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と雑音の基底スペクトル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ことが可能となる、という原理に基づく。従って、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には推定したスペクトルが実際の音声スペクトルに対応しない可能性がある。このため、音声スペクトルと雑音スペクトルの分解の不定性を解消するためには音声スペクトルが満たすべき

より強い制約が必要である。また、SSNMF 法では信号は強調できたとしても特徴量を強調できる保証はないため、強調処理が音声認識や音声変換など音声特徴量に基づく音声処理の性能向上に直結するとは限らない。そこで本稿では、スペクトルだけでなく特徴量の事前情報も活用することで音声スペクトルを推定する手がかりをより多く与えるとともに特徴量の歪みを生じにくくする SSNMF の正則化法を提案する。

2 問題の定式化

観測信号の振幅スペクトログラムまたはパワースペクトログラム（以後、観測スペクトログラム）を $Y_{\omega,t}$ とする。ただし、 ω と t は周波数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 $X_{\omega,t}^{(s)}$ および雑音スペクトル $X_{\omega,t}^{(n)}$ をそれぞれ K_s 個の基底スペクトル $H_{1,\omega}^{(s)}, \dots, H_{K_s,\omega}^{(s)}$ と K_n 個の基底スペクトル $H_{1,\omega}^{(n)}, \dots, H_{K_n,\omega}^{(n)}$ の非負結合

$$X_{\omega,t}^{(s)} = \sum_{k=1}^{K_s} H_{k,\omega}^{(s)} U_{k,t}^{(s)} \quad (1)$$

$$X_{\omega,t}^{(n)} = \sum_{k=1}^{K_n} H_{k,\omega}^{(n)} U_{k,t}^{(n)} \quad (2)$$

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 $H_{1,\omega}^{(s)}, \dots, H_{K_s,\omega}^{(s)}$ を用いて、観測スペクトル $Y_{\omega,t}$ に $X_{\omega,t} = X_{\omega,t}^{(s)} + X_{\omega,t}^{(n)}$ 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。このようにして求まる音声スペクトルと雑音スペクトルの推定値から Wiener フィルタなどにより観測信号から音声信号を得ることができる。このアプローチでは事前学習した音声の基底スペクトルが音声と雑音の分離の手がかりとなるが、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 $X_{\omega,t}^{(s)}$ と $X_{\omega,t}^{(n)}$ 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。このため、同じ $X_{\omega,t}$ を与える $X_{\omega,t}^{(s)}$ と $X_{\omega,t}^{(n)}$ の不定性を解消するためには音声スペクトルが満たすべきより強い制約が必要である。今、もし $X_{\omega,t}^{(s)}$ が音声スペクトルに対応しているなら $X_{\omega,t}^{(s)}$ は特徴量空間においても音声が実際にとりうる範囲内に分布するはずである。そこで、本稿では、ケプストラム特徴量（後述）に着目し、ケプストラム空間で定義される確率分布に基づいて $X_{\omega,t}^{(s)}$ に対する正則化項を考え、これと $Y_{\omega,t}$ と $X_{\omega,t}$ の誤差規準の和を規準としたパラメータ最適化アルゴリズムを提案する。

*Speech Enhancement Based on Semi-Supervised Non-negative Matrix Factorization with Cepstral Distance Regularization, Li Li (The University of Tokyo), Hirokazu Kameoka (NTT Communication Science Laboratories/The University of Tokyo), Takuya Higuchi (NTT Communication Science Laboratories), Hiroshi Saruwatari (The University of Tokyo)

$Y_{\omega,t}$ と $X_{\omega,t}$ の誤差は二乗誤差, I ダイバージェンス, 板倉齋藤距離などで測ることができるが, ここでは I ダイバージェンス

$$\mathcal{I}(Y|X) = \sum_{\omega,t} \left(Y_{\omega,t} \log \frac{Y_{\omega,t}}{X_{\omega,t}} - Y_{\omega,t} + X_{\omega,t} \right) \quad (3)$$

を用いる。ただし, すべての基底スペクトルは

$$\sum_{\omega} H_{k,\omega}^{(s)} = 1, \quad \sum_{\omega} H_{k,\omega}^{(n)} = 1 \quad (4)$$

のような制約を満たしているものとする。次に, $X_{\omega,t}^{(s)}$ に対し,

$$\mathcal{K}(X^{(s)}) = \log \prod_t \sum_m w_m \prod_n \mathcal{N}(\mathcal{X}_{n,t}; \mu_{n,m}, \sigma_{n,m}^2) \quad (5)$$

$$\mathcal{X}_{n,t} = \sum_l c_{n,l} \log \sum_{\omega} f_{l,\omega} X_{\omega,t}^{(s)} \quad (6)$$

のような規準を考える。ただし, $\mathcal{X}_t = (\mathcal{X}_{0,t}, \dots, \mathcal{X}_{N-1,t})^T$ は $X_{0,t}, \dots, X_{\Omega-1,t}$ のメル周波数ケプストラム係数 (Mel-Frequency Cepstrum Coefficients: MFCC) であり, $f_{l,\omega}$ は l 番目のメルフィルタバンク係数, $\{c_{n,l}\}_{0 \leq n \leq N-1, 0 \leq l \leq N-1}$ は離散コサイン変換の係数である。式 (5) は, $\mathcal{X}_0, \dots, \mathcal{X}_{T-1}$ がパラメータ $\theta = \{\mu_m, \Sigma_m, w_m\}_{1 \leq m \leq M}$ の混合正規分布から生成される確率の対数を表す。ただし, $\mu_m = (\mu_{1,m}, \dots, \mu_{N,m})^T$, $\Sigma_m = \text{diag}(\sigma_{1,m}, \dots, \sigma_{N,m})$, w_m は m 番目の正規分布の平均と分散と重みを表す。クリーン音声の学習サンプルの MFCC 系列からこの混合正規分布のパラメータ θ を学習することで, $\mathcal{K}(X^{(s)})$ を, $X_{\omega,t}^{(s)}$ が MFCC 空間においてできるだけ学習サンプルと同様に分布する場合に高いスコアを与える規準とすることができる。提案法は, 式 (3) と式 (5) の二つの規準を考慮した

$$\mathcal{J}(U^{(s)}, H^{(n)}, U^{(n)}) = \mathcal{I}(Y|X) - \lambda \mathcal{K}(X^{(s)}) \quad (7)$$

のような規準を最小化することが目的である。ただし, λ は正則化パラメータである。以上のようにこの最適化問題はスペクトルのモデルをケプストラム距離規準でソフトな制約を課す問題となっており, これまで我々はこの枠組により楽音分離と音色クラスタリングを同一最適化規準の下で行う手法を提案している [9, 10]。提案法はこの枠組により音声の信号強調と特徴量強調を同時に実現することを目指した手法であり, 「ケプストラム正則化 SSNMF」と呼ぶ。

3 パラメータ推定アルゴリズム

$\mathcal{J}(U^{(s)}, H^{(n)}, U^{(n)})$ を最小化する $U^{(s)}, H^{(n)}, U^{(n)}$ を解析的に得ることはできないが, [9, 10] と同様に, 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 α を導入し, $F(\theta) = \min_{\alpha} F^+(\theta, \alpha)$ を満たす補助関数 $F^+(\theta, \alpha)$ を設計する。このような補助関数が設計できれば, $\alpha \leftarrow \text{argmin}_{\alpha} F^+(\theta, \alpha)$ と $\theta \leftarrow \text{argmin}_{\theta} F^+(\theta, \alpha)$ 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。以下で, $\mathcal{J}(U^{(s)}, H^{(n)}, U^{(n)})$ の補助関数とそれに基づく更新式を導く。

$\mathcal{I}(Y|X)$ については, 負の対数関数が凸関数であることを利用し, Jensen の不等式により

$$\mathcal{I}(Y|X) \leq \mathcal{I}^+(Y|X) \quad (8)$$

$$\mathcal{I}^+(Y|X) \stackrel{c}{=} \sum_{\omega,t} \left(-Y_{\omega,t} \sum_k \zeta_{k,\omega,t} \log \frac{H_{k,\omega} U_{k,t}}{\zeta_{k,\omega,t}} + X_{\omega,t} \right)$$

のような上界関数が立てられる。ただし, $\stackrel{c}{=}$ はパラメータに依存する項のみに関する等号を表す。また H と U は, ここでは

$$H_{k,\omega} = H_{k,\omega}^{(s)} \quad (k = 1, \dots, K_s) \quad (9)$$

$$H_{k+K_s,\omega} = H_{k,\omega}^{(n)} \quad (k = 1, \dots, K_n) \quad (10)$$

$$U_{k,t} = U_{k,t}^{(s)} \quad (k = 1, \dots, K_s) \quad (11)$$

$$U_{k+K_s,t} = U_{k,t}^{(n)} \quad (k = 1, \dots, K_n) \quad (12)$$

としている。 $\zeta_{k,\omega,t}$ は $\zeta_{k,\omega,t} \geq 0$, $\sum_k \zeta_{k,\omega,t} = 1$ を満たす変数であり, 式 (8) の等号は

$$\zeta_{k,\omega,t} = \frac{H_{k,\omega} U_{k,t}}{\sum_{k'} H_{k',\omega} U_{k',t}} \quad (13)$$

のとき成立する。

次に, $-\mathcal{K}(X^{(s)})$ の上界関数を設計する。式 (8) と同様, 負の対数関数が凸関数であることを利用し, Jensen の不等式より

$$\begin{aligned} -\mathcal{K}(X^{(s)}) &\leq -\sum_{t,m} \alpha_{m,t} \log \frac{w_m \prod_n \mathcal{N}(\mathcal{X}_{n,t}; \mu_{n,m}, \sigma_{n,m}^2)}{\alpha_{m,t}} \\ &\stackrel{c}{=} \sum_{t,m} \alpha_{m,t} \sum_n \frac{(\mathcal{X}_{n,t} - \mu_{n,m})^2}{2\sigma_{n,m}^2} \end{aligned} \quad (14)$$

のような不等式が立てられる。式 (14) の等号は

$$\alpha_{m,t} = \frac{w_m \prod_n \mathcal{N}(\mathcal{X}_{n,t}; \mu_{n,m}, \sigma_{n,m}^2)}{\sum_{m'} w_{m'} \prod_n \mathcal{N}(\mathcal{X}_{n,t}; \mu_{n,m'}, \sigma_{n,m'}^2)} \quad (15)$$

のとき成立する。続いて $(\mathcal{X}_{n,t} - \mu_{n,m})^2$ の上界関数を導く。二次関数は凸関数なので, Jensen の不等式より

$$(\mathcal{X}_{n,t} - \mu_{n,m})^2 \leq \sum_l \frac{(c_{n,l} \log G_{l,t} - \varphi_{l,n,m,t})^2}{\beta_{l,n,m,t}} \quad (16)$$

のような不等式が立てられる。ただし, $G_{l,t} = \sum_{\omega} f_{l,\omega} X_{\omega,t}^{(s)}$ である。 $\beta_{l,n,m,t}$ は $\sum_l \beta_{l,n,m,t} = 1$ を満たす任意の正の定数, $\varphi_{l,n,m,t}$ は $\sum_l \varphi_{l,n,m,t} = \mu_{n,m}$ を満たす変数であり, 式 (16) の等号は

$$\varphi_{l,n,m,t} = c_{n,l} \log G_{l,t} + \beta_{l,n,m,t} (\mu_{n,m} - \mathcal{X}_{n,t}) \quad (17)$$

のとき成立する。式 (16) と式 (14) より,

$$-\mathcal{K}(X^{(s)}) \leq \sum_{t,l} A_{l,t} (\log G_{l,t})^2 + \sum_{t,l} B_{l,t} \log G_{l,t} + d$$

$$\begin{aligned} d &= -\sum_{t,m} \alpha_{m,t} \log \frac{w_m}{\alpha_{m,t}} \\ &+ \frac{1}{2} \sum_{t,m} \alpha_{m,t} \sum_n \log 2\pi\sigma_{n,m}^2 + \sum_{t,n,m,l} \frac{\alpha_{m,t} \varphi_{l,n,m,t}^2}{2\sigma_{n,m}^2 \beta_{l,n,m,t}} \end{aligned}$$

がいえる。ただし、

$$G_{l,t} = \sum_{\omega} f_{l,\omega} X_{\omega,t}^{(s)} \quad (18)$$

$$A_{l,t} = \sum_{n,m} \frac{\alpha_{m,t} c_{n,l}^2}{2\sigma_{n,m}^2 \beta_{l,n,m,t}} \quad (19)$$

$$B_{l,t} = - \sum_{n,m} \frac{\alpha_{m,t} c_{n,l} \varphi_{l,m,n,t}}{\sigma_{n,m}^2 \beta_{l,n,m,t}} \quad (20)$$

である。 $A_{l,t}$ は非負値である点に注意し、次に $(\log G_{l,t})^2$ の上界関数を考える。 $(\log G_{l,t})^2$ の上界は [9] の不等式を用いて

$$(\log G_{l,t})^2 \leq \frac{1}{G_{l,t}} + p(\xi_{l,t})G_{l,t} + q(\xi_{l,t}) \quad (21)$$

で与えることができる。ただし、

$$p(\xi_{l,t}) = \frac{2 \log \xi_{l,t}}{\xi_{l,t}} + \frac{1}{\xi_{l,t}^2} \quad (22)$$

$$q(\xi_{l,t}) = (\log \xi_{l,t})^2 - 2 \log \xi_{l,t} - \frac{2}{\xi_{l,t}} \quad (23)$$

であり、式 (21) の等号は

$$\xi_{l,t} = G_{l,t} \quad (24)$$

のとき成立する。さらに、 $f_{l,\omega} H_{k,\omega} U_{k,t}$ が非負値であること、逆数関数が正領域で凸関数であることから、Jensen の不等式より

$$\frac{1}{G_{l,t}} = \frac{1}{\sum_{\omega,k} f_{l,\omega} H_{k,\omega} U_{k,t}} \quad (25)$$

$$\leq \sum_{\omega,k} \frac{\rho_{l,k,\omega,t}^2}{f_{l,\omega} H_{k,\omega} U_{k,t}} \quad (26)$$

が成り立つ。ただし、 $\rho_{l,k,\omega,t}$ は $\rho_{l,k,\omega,t} > 0$ 、 $\sum_{\omega,k} \rho_{l,k,\omega,t} = 1$ を満たす変数であり、

$$\rho_{l,k,\omega,t} = \frac{f_{l,\omega} H_{k,\omega} U_{k,t}}{\sum_{\omega',k'} f_{l,\omega'} H_{k',\omega'} U_{k',t}} \quad (27)$$

のとき式 (26) の等号は成立する。続いて $B_{l,t} \log G_{l,t}$ の項の上界を考える。 $B_{l,t}$ は非負値であるとは限らないので、 $B_{l,t}$ の符号に応じて別種の不等式を立てる。まず、対数関数が凹関数であるため、 $B_{l,t} \geq 0$ のとき、

$$B_{l,t} \log G_{l,t} \leq B_{l,t} \left(\frac{G_{l,t}}{\phi_{l,t}} + \log \phi_{l,t} - 1 \right) \quad (28)$$

のような不等式を得る。 $\phi_{l,t}$ は正の変数であり、

$$\phi_{l,t} = G_{l,t} \quad (29)$$

のとき式 (28) の等号は成立する。一方 $B_{l,t} < 0$ のとき、負の対数関数は凸関数より Jensen の不等式により

$$B_{l,t} \log G_{l,t} \leq B_{l,t} \sum_{\omega,k} v_{k,l,\omega,t} \log \frac{f_{l,\omega} H_{k,\omega} U_{k,t}}{v_{k,l,\omega,t}} \quad (30)$$

がいえる。ただし、 $v_{k,l,\omega,t}$ は $v_{k,l,\omega,t} > 0$ 、 $\sum_{k,\omega} v_{k,l,\omega,t} = 1$ を満たす変数であり、

$$v_{k,l,\omega,t} = \frac{f_{l,\omega} H_{k,\omega} U_{k,t}}{\sum_{\omega',k'} f_{l,\omega'} H_{k',\omega'} U_{k',t}} \quad (31)$$

のとき式 (30) の等号は成立する。まとめると、

$$B_{l,t} \log G_{l,t} \leq \delta_{B_{l,t} \geq 0} |B_{l,t}| \left(\frac{G_{l,t}}{\phi_{l,t}} + \log \phi_{l,t} - 1 \right) - \delta_{B_{l,t} < 0} |B_{l,t}| \sum_{\omega,k} v_{k,l,\omega,t} \log \frac{f_{l,\omega} H_{k,\omega} U_{k,t}}{v_{k,l,\omega,t}}$$

と書ける。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。以上より、 $-\mathcal{K}(X^{(s)})$ の上界関数

$$\begin{aligned} & -\mathcal{K}(X^{(s)}) \quad (32) \\ & \leq \sum_{t,l} A_{l,t} \left(\sum_{\omega,k} \frac{\rho_{l,k,\omega,t}^2}{f_{l,\omega} H_{k,\omega} U_{k,t}} + p(\xi_{l,t})G_{l,t} + q(\xi_{l,t}) \right) \\ & + \sum_{t,l} \delta_{B_{l,t} \geq 0} |B_{l,t}| \left(\frac{G_{l,t}}{\phi_{l,t}} + \log \phi_{l,t} - 1 \right) \\ & - \sum_{t,l} \delta_{B_{l,t} < 0} |B_{l,t}| \sum_{\omega,k} v_{k,l,\omega,t} \log \frac{f_{l,\omega} H_{k,\omega} U_{k,t}}{v_{k,l,\omega,t}} + d \end{aligned}$$

を得ることができる。この不等式を導いたことのポイントは、右辺を最小にする H や U を解析的に得ることができる点にあり、 \mathcal{I}^+ と合わせるにより更新式を閉形式で与える補助関数を設計することができる。この補助関数より、各パラメータの更新式

$$H_{k,\omega}^{(s)} = \frac{-b_{k,\omega} + \sqrt{b_{k,\omega}^2 - 4a_{k,\omega}c_{k,\omega}}}{2a_{k,\omega}} \quad (33)$$

$$H_{k,\omega}^{(n)} = \frac{\sum_t \zeta_{k+K_s,\omega,t} Y_{\omega,t}}{\sum_t U_{k+K_s,t}} \quad (34)$$

$$U_{k,t}^{(s)} = \frac{-e_{k,t} + \sqrt{e_{k,t}^2 - 4d_{k,t}f_{k,t}}}{2d_{k,t}} \quad (35)$$

$$U_{k,t}^{(n)} = \frac{\sum_{\omega} \zeta_{k+K_s,\omega,t} Y_{\omega,t}}{\sum_{\omega} H_{k+K_s,\omega}} \quad (36)$$

を得る。ただし、

$$a_{k,\omega} = \sum_t U_{k,t} + \lambda \sum_{l,t} A_{l,t} p(\xi_{l,t}) f_{l,\omega} U_{k,t} + \lambda \sum_{l,t} \frac{\delta_{B_{l,t} \geq 0} |B_{l,t}|}{\phi_{l,t}} f_{l,\omega} U_{k,t}$$

$$b_{k,\omega} = - \sum_t \zeta_{k,\omega,t} Y_{\omega,t} - \lambda \sum_{l,t} \delta_{B_{l,t} < 0} |B_{l,t}| v_{k,l,\omega,t}$$

$$c_{k,\omega} = -\lambda \sum_{l,t} \frac{A_{l,t} \rho_{l,k,\omega,t}^2}{f_{l,\omega} U_{k,t}}$$

$$d_{k,t} = \sum_{\omega} H_{k,\omega} + \lambda \sum_{\omega,l} A_{l,t} p(\xi_{l,t}) f_{l,\omega} H_{k,\omega} + \lambda \sum_{\omega,l} \frac{\delta_{B_{l,t} \geq 0} |B_{l,t}|}{\phi_{l,t}} f_{l,\omega} H_{k,\omega}$$

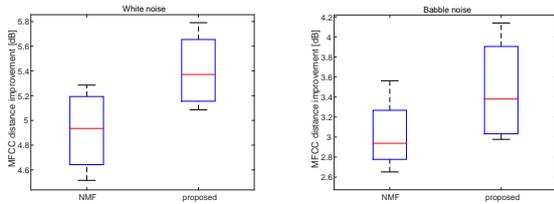


Fig. 1 提案法と従来法によって得られたケプストラム歪みの改善値。雑音が白色雑音の場合 (左) とバブル雑音の場合 (右)。

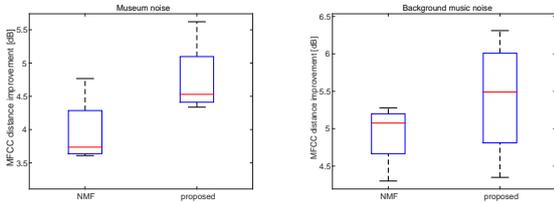


Fig. 2 提案法と従来法によって得られたケプストラム歪みの改善値。雑音が実環境雑音の場合 (左) と背景音楽の場合 (右)。

$$e_{k,t} = -\sum_{\omega} \zeta_{k,\omega,t} Y_{\omega,t} - \lambda \sum_{\omega,l} \delta_{B_{l,t} < 0} |B_{l,t}| v_{k,l,\omega,t}$$

$$f_{k,t} = -\lambda \sum_{\omega,t} \frac{A_{l,t} \rho_{l,k,\omega,t}^2}{f_{l,\omega} H_{k,\omega}}$$

である。

4 評価実験

ATR 音声データベース 503 文の音声データと RWCP の雑音データ (white noise, babble noise, museum noise, background music noise の 4 種類) 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、処理前と処理後の信号対歪み比 (SDR) およびケプストラム歪みの改善値を評価した。テストデータはクリーン音声に各雑音をさまざまな SNR で重畳させて作成した。テストデータはすべて音響信号はサンプリング周波数 16kHz のモノラル信号で、フレーム長 32ms, フレームシフト 16ms 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 10 名 (うち女性 4 名, 男性 6 名) の話者の計 450 文の音声を用いて $H_{k,\omega}^{(s)}$ と MFCC の GMM パラメータ θ の学習を行った。MFCC の次元は 13 とし GMM の混合数は 30 とした。テストにおいては、学習で得られた $H_{k,\omega}^{(s)}$ と θ を固定し、 $\lambda = 1$ とし $U_{k,t}^{(s)}$, $H_{k,\omega}^{(n)}$, $U_{k,t}^{(n)}$ の推定を行った。推定後、 $X_{\omega,t}^{(s)}$ と $X_{\omega,t}^{(n)}$ 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。提案法アルゴリズムの初期値は従来の SSNMF により得た。

以上の条件下での提案法と従来法によって得られたケプストラム歪みおよび SDR の改善値を図 1~4 に示す。いずれの評価尺度においてもほとんどの場合において提案法の方が高い改善値を得られていることが確認できる。

5 おわりに

SSNMF による音声強調では、雑音とクリーン音声で類似するスペクトルが存在する場合、音声の基底スペクトルで雑音スペクトルを表現してしまうこと (あ

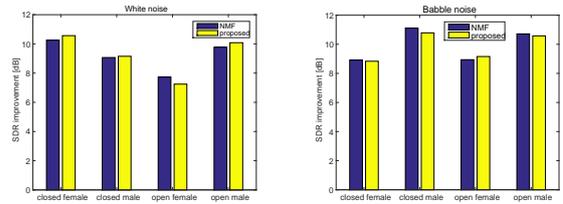


Fig. 3 提案法と従来法によって得られた SDR の改善値。雑音が白色雑音の場合 (左) とバブル雑音の場合 (右)。

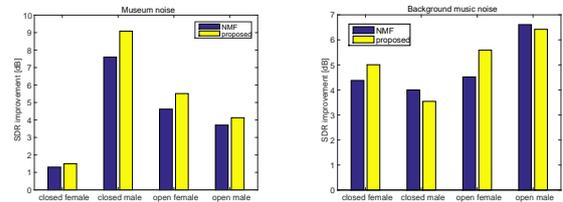


Fig. 4 提案法と従来法によって得られた SDR の改善値。雑音の種類が実環境雑音の場合 (左) と背景音楽雑音の場合 (右)。

るいはその逆) により性能が低下することがあった。また、この手法では信号は強調できたとしても特徴量を強調できる保証はなく、逆に特徴量を損傷してしまう場合がある。このため強調処理が音声認識や音声変換などのような音声特徴量に基づく音声処理の性能向上に必ずしも直結しないという問題があった。そこで本稿ではこの問題を解決するため、音声の MFCC 空間の確率分布を正則化規準とした SSNMF による音声強調法「ケプストラム距離正則化 SSNMF」を提案し、実験によりその効果を確認した。

謝辞 本研究は JSPS 科研費 26730100 の助成を受けて行われた。

参考文献

- [1] P.C. Loizou, *Speech Enhancement: Theory and Practice*, Boca Raton, FL: CRC Press, 2007.
- [2] J. Moreno et al., "A vector Taylor series approach for environment-independent speech recognition," in Proc. ICASSP, 2, 733-736, 1996.
- [3] T. Kristjansson and J. Hershey "High resolution signal reconstruction," in Proc. ASRU, 291-296, 2003.
- [4] J. Droppo et al., "Evaluation of SPLICE on the Aurora 2 and 3 tasks," in Proc. ICSLP, 29-32, 2002.
- [5] J. Droppo and A. Acero, "Maximum mutual information SPLICE transform for seen and unseen conditions," in Proc. Interspeech, 989-992, 2005.
- [6] Y. Xu et al., "An experimental study on speech enhancement based on deep neural networks," IEEE SPL, 21(1), 65-68, 2014.
- [7] M. Kim and P. Smaragdis, "Adaptive denoising autoencoders: A fine-tuning scheme to learn from test mixtures," in Proc. LVA/ICA, 2015.
- [8] P. Smaragdis et al., "Supervised and semi-supervised separation of sounds from single-channel mixtures," in Proc. ICA, 414-421, 2007.
- [9] H. Kameoka et al., "Constrained and regularized variants of non-negative matrix factorization incorporating music-specific constraints," in Proc. ICASSP, 5365-5368, 2012.
- [10] 樋口, 亀岡, "ケプストラム距離正則化に基づく多重音解析," 情報研報, 2014-MUS-104(10), 2014.