On Rapid Adaptation of Noise Spectral Estimation in Spectral Suppression Method

第22回 信号処理シンポジウム
2007年11月7日〜9日（仙台）

スペクトルサプレッション法における雑音スペクトル推定法の検討 -急激な時間的変化への対応-

On Rapid Adaptation of Noise Spectral Estimation in Spectral Suppression Method

東 尚哉 † 中山 謙二 † 平野 晃宏 †
† 金沢大学大学院 自然科学研究科 電子情報工学専攻

Shoya Higashi† Kenji Nakayama† Akihiro Hirano†
†Division of Electronics and Computer Sicience
Graduate School of Natural Science and Technology, Kanazawa Univ.
E-mail : higashi@leo.ec.t.kanazawa-u.ac.jp
nakayama@t.kanazawa-u.ac.jp

あらまし

本稿では，スペクトルサプレッション法を用いた単一マイク方式のノイズキャンセラにおける，急激な雑音スペクトルの変化に追従できる雑音スペクトル推定について検討する．無音区間と有音区間を各フレームのスペクトルエントロピーを用いて検出する Voice Activity Detector によって検出し，無音区間ではパラメータを適宜制御したリーク積分による雑音スペクトル推定，音声区間では雑音抑圧精度が高い重み付き雑音推定により雑音スペクトルを推定している．この方法により，従来法に比べて雑音スペクトルの急変化にも追従することができ，また正規化推定誤差と SNR も従来法より向上した．

ABSTRACT

A noise spectral estimator in a spectral suppression method is proposed. Especially, a rapid adaptation for noise spectral change is taken into account. In order to estimate the noise spectrum quickly and accurately, a detection method for a speech-absent frame and a speech-present frame by using a voice activity detector (VAD) is improved. Furthermore, an improved noise spectral estimation method for the speech-absent frame is proposed. The conventional method is applied to the speech-present frame. The proposed method can estimate the noise spectrum more precisely than the conventional methods. The segmental SNR is improved by 2.0〜3.8 dB and the normalized estimated error is improved by about 3.2〜4.7 dB for white noise and babble noise, which are combined.

1 まえがき

現在，携帯電話などの移動通信が普及し，街頭や車内など背景雑音が多い場所で携帯電話が使用される場合も多い．このような雑音を除去するための方法として，単一マイク方式のノイズキャンセラが開発されている．

単一マイク方式のノイズキャンセラとして，スペクトルサプレッション法が研究されている [1]-[4],[9]．この方式は雑音混入信号のスペクトルと雑音スペクトルの比を雑音混入信号に乗じて雑音成分を抑制する方法である．雑音の抑圧度を決めるスペクトルゲインを求める方法として MMSE STSA 法 [1] や Joint MAP 法 [2] がある．

スペクトルサプレッション法において，スペクトルゲインを計算するためには雑音スペクトルが必要である．雑音スペクトル推定が不正確だと，雑音抑圧後に雑音が大きく残ったり，雑音の過大推定により雑音抑圧後に音声が大きく歪み，音質が劣化する．以上のことより，雑音スペクトルをいかに正確に求められるかが重要である．

従来法では，雑音スペクトルは約 20 フレームに亘る平均値として推定されており，雑音の急激な変化に追従することが難しい．これに対して，雑音の急激な変化に適応できる方式 (Rapid Adaptation) も研究されている [6]．本稿では，この Rapid Adaptation の方式に基づいて，雑音の急激な変化に追従し，より正確に雑音スペクトルを推定する方法を提案する．

2 スペクトルサプレッション法

2.1 スペクトルサプレッション法の構成

図 1 にスペクトルサプレッション法の構成図を示す．音声と雑音はともにスペクトル成分において統計的独立であるとする．時間領域でのクリア音声を $x(n)$，雑音を $d(n)$ とおくと，雑音混入音声 $x(n)$ は，

$$x(n) = s(n) + d(n) \tag{1}$$

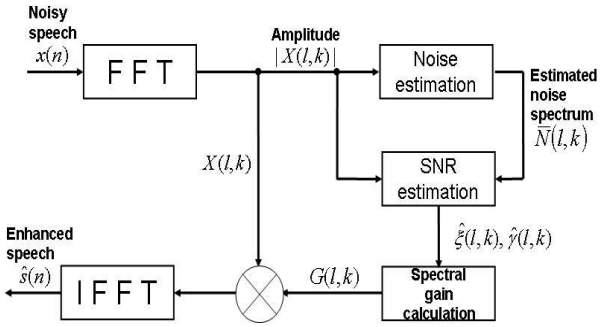

図 1: スペクトルサプレッション法のブロック図

と表せる．音声信号は一般に非定常でありその音響的特徴は変動している．そのため，音声のスペクトル分析では十分に短い時間の区間において音声は定常状態であるという仮定の基で，少しずつ時間区間をシフトさせながら窓関数を用いて切り出したフレームの波形のデータに対して順次 FFT 演算を実行し，スペクトル時系列を得ている．よって雑音混入音声は M サンプルのフレームに分けられていて，2M サンプルの窓関数を用いて 50% オーバーラップさせることにより，n 番目のフレームおける切り出された雑音混入音声 $\hat{x}_n(n)$ は次の式のように表せる．

$$\hat{x}_n(n) = \begin{cases} h(n)x_{n-1}(n) & ,1 \le n \le M \\ h(n)x_n(n-M) & ,M \le n \le 2M \end{cases} \quad (2)$$

この信号の l 番目のフレームにおける k 番目の周波数領域での表示を次のように表す.

$$X(l,k) = S(l,k) + D(l,k) \quad (3)$$

事前 SNR(クリーン音声対雑音比)，事後 SNR(雑音混入音声対雑音比) はそれぞれ次の式で表せる．

$$\xi(l,k) = \frac{E\{|S(l,k)|^2\}}{E\{|D(l,k)|^2\}} \quad (4)$$

$$\gamma(l,k) = \frac{|X(l,k)|^2}{E\{|D(l,k)|^2\}} \quad (5)$$

実際に利用可能なものは，雑音混入音声のみで，事前 SNR, 事後 SNR は推定しなくてはいけない．事前 SNR は decision-directed 方式で以下のように推定できる [1].

$$\hat{\xi}(l,k) = \alpha\gamma(l-1,k)G^2(l-1,k)+(1-\alpha)P[\gamma(l,k)-1] \quad (6)$$

ただし，α は $0 < \alpha < 1$，$P[x]$ は次の式を満たす．

$$P[x] = \begin{cases} x & (x > 0) \\ 0 & (otherwise) \end{cases} \quad (7)$$

事後 SNR は推定した雑音スペクトル $N(l,k)$ を用いて，次のように推定する．

$$\hat{\gamma}(l,k) = \frac{|X(l,k)|^2}{N(l-1,k)} \quad (8)$$

以上のように推定した事前 SNR，事後 SNR によりスペクトルゲイン $G(l,k)$ を求め，雑音混入音声に乗じることで雑音を抑える．

2.2 MMSE STSA 法

MMSE STSA 法は雑音混入音声から正確な音声のスペクトル振幅を抽出し，その二乗誤差を最小にする方式である [1]．統計モデルとして，音声と雑音ともにスペクトル成分において統計的独立で平均 0 のガウス分布であると仮定する．MMSE STSA 法におけるスペクトルゲインは，

$$G(l,k) = \left[(1+\nu(l,k))I_0\left(\frac{\nu(l,k)}{2}\right) + \nu(l,k)I_1\left(\frac{\nu(l,k)}{2}\right)\right] \cdot \frac{\Lambda(l,k)}{1+\Lambda(l,k)} \frac{\sqrt{\pi\nu(l,k)}}{2\gamma(l,k)} \exp\left(-\frac{\nu(l,k)}{2}\right) \quad (9)$$

で求められる．式中の各関数は

$$\nu(l,k) = \frac{\eta(l,k)}{1+\eta(l,k)} \cdot \hat{\gamma}(l,k) \quad (10)$$

$$\Lambda(l,k) = \frac{1-q(l,k)}{q(l,k)} \cdot \frac{\exp(\nu(l,k))}{1+\nu(l,k)} \quad (11)$$

$$\eta(l,k) = \frac{\hat{\xi}(l,k)}{1-q(l,k)} \quad (12)$$

$$q(l,k) = \alpha_q q(l-1,k) + (1-\alpha_q)I(l,k) \quad (13)$$

で求められる．I_0,I_1 はそれぞれ 0 次と 1 次の Bessel 関数である．$q(l,k)$ は事前確率と言い，非音声成分及びパワーが十分に小さい音声成分が，雑音混入音声に含まれる確率を表している [5]．ただし，$\gamma(m,k)$ が閾値 γ_{th} よりも大きい時は，$I(m,k) = 1$ とし，$\gamma(m,k)$ が γ_{th} よりも小さい時は $I(m,k) = 0$ とする．

2.3 Joint MAP 法

Joint MAP 法は，雑音はガウス分布，音声をスーパーガウス分布という仮定のもとでスペクトルゲインを計算する方法である [2].

Joint MAP 法におけるスペクトルゲインは，

$$G(l,k) = u(l,k) + \sqrt{u^2(l,k) + \frac{\tau}{2\hat{\gamma}(l,k)}} \quad (14)$$

$$u(l,k) = \frac{1}{2} - \frac{\mu}{4\sqrt{\hat{\gamma}(l,k)\hat{\xi}(l,k)}} \quad (15)$$

と求められる．

2.4 スペクトルゲイン補正

無音区間のフレームにおいて，スペクトルゲインを更に抑圧し，より雑音を抑えた雑音抑圧音声とするため，スペクトルゲイン $G(m,k)$ に次に示す倍率 G_{sup} を乗じる．

$$G(m,k) = \begin{cases} G_{sup}G(m,k), & (\text{無音区間}) \\ G(m,k), & (\text{有音区間}) \end{cases} \quad (16)$$

次に，過剰抑圧による音質の劣化を避けるために，スペクトルゲインの最小値の制限 [3],[8] と，SNR に基づいて原音 (観測信号 = 音声 + 雑音) をある割合だけ付加した [9]．次式のように，スペクトルゲインの最小値を G_{floor} により制限する．

$$G(m,k) = \begin{cases} G(m,k), & G(m,k) > G_{floor} \\ G_{floor}, & G(m,k) \leq G_{floor} \end{cases} \quad (17)$$

ただし，G_{floor} の大きさは，有音区間では無音区間における G_{floor} よりも大きく設定している [8]．

3 従来の高速追従形雑音スペクトル推定法

本節では，音声区間検出法：Voice Activity Detector (VAD) を用いた従来の高速追従形雑音スペクトル推定法について述べる [6]．

3.1 VAD -Voice Activity Detector-

Voice Activity Detector（以下，VAD とする）は，入力信号のスペクトルエントロピー $H(l)$ を用いた音声区間検出である [7]．無音区間では，スペクトルエントロピーは音声フレームに比べて大きくなる．そこで，入力信号の最初の区間を無音区間と仮定し，最初の数フレーム分のスペクトルエントロピーの平均値に定数 c を掛けたものを閾値 σ とし，その後のフレームでは，スペクトルエントロピーが閾値よりも小さい場合は音声区間，閾値よりも大きい場合は無音区間とする．スペクトルエントロピー $H(l)$ は次のように求められる．

$$H(l) = -\sum_{k=1}^{2M} P_r(l,k) \cdot \log(P_r(l,k)) \quad (18)$$

$$P_r(l,k) = \frac{X_{energy}(l,k)}{\sum_{k=1}^{2M} X_{energy}(l,k)} \quad (19)$$

$$X_{energy}(l,k) = |X(l,k)|^2 \quad (20)$$

ただし，式中の $2M$ は周波数のデータ数である．また，音声スペクトルのほとんどが周波数帯域 $250Hz$ 以上，$4000Hz$ 以下に存在するので，次のように定める．

$$X_{energy}(l,k) = 0, \quad k \leq 250 \ or \ k \geq 4000 \quad (21)$$

また，論文 [7] より，スペクトルエントロピーについて次のように報告されている．

- 式 (19) より，スペクトルエントロピーは正規化されているので，スペクトルの分布が変化しない限り，音声スペクトルの大きさが変化しても，理論的にはスペクトルエントロピーは変化しない．しかし，音声区間と無音区間でのエントロピーの大きさの違いは，音声スペクトルが小さい時に小さくなる．
- 大半の雑音スペクトルは，音声スペクトルと異なった確率分布となるため，音声と雑音のスペクトルエントロピーは異なる．
- スペクトルエントロピーは，雑音にロバストである．ただし，SNR が低い場合，音声区間と無音区間でのエントロピーの大きさの違いは小さくなるので，無音区間の検出が難しくなる．

このように，雑音が大きい場合や，音声スペクトルが小さい場合，音声区間と無音区間でのエントロピーの変化が小さくなるため，無音フレームの検出が難しくなる．そこで，式 (19) において，正の定数 C を加えることで，雑音が大きい場合でも音声スペクトルが小さい場合でも，音声区間と無音区間のエントロピーの変化を大きくすることで，VAD の精度を高めることができる [7]．

$$\acute{H}(l) = -\sum_{k=1}^{2M} \acute{P}_r(l,k) \cdot \log(\acute{P}_r(l,k)) \quad (22)$$

$$\acute{P}_r(l,k) = \frac{X_{energy}(l,k) + C}{\sum_{k=1}^{2M} X_{energy}(l,k) + C} \quad (23)$$

3.2 Rapid Adaptation

Rapid Adaptation について説明する．Rapid Adaptation とは，VAD を用いて音声フレームか無音フレームかを判断し，そのフレームに適した雑音推定アルゴリズムを適用することで，急激に雑音環境が変化した場合でも，高速かつ正確に雑音スペクトルを推定するアルゴリズムである [6]．以下に無音フレームと音声フレームで用いるアルゴリズムについて説明する．

3.2.1 無音フレーム

VAD において，入力信号のスペクトルエントロピー $H(l)$ が閾値 σ より大きくなったとき，そのフレームは無音フレームと判断する．無音フレームにおける雑音混入音声のスペクトルは，雑音スペクトルに等しいので，無音フレームをトラッキングすることで雑音推定スペクトルを次のように更新する．

$$\bar{N}(l,k) = \lambda \cdot \bar{N}(l-1,k) + (1-\lambda) \cdot |X(l,k)|^2 \quad (24)$$

しかしこの方法では，パラメータ λ が定数であるため，時間とともに変化する雑音スペクトルには対応できない．

3.2.2 音声フレーム

VAD において，入力信号のスペクトルエントロピー $H(l)$ が閾値 σ より小さくなったとき，そのフレームは音声フレームと判断する．従来法では，Cohen のアルゴリズム [10] を用いて式 (25) のように平滑化パラメータを適応的に制御しながら雑音スペクトルを推定している．

$$\bar{N}(l,k) = \rho(l,k) \cdot \bar{N}(l-1,k) + (1-\rho(l,k)) \cdot |X(l,k)|^2 \quad (25)$$

$$\rho(l,k) = a_d + (1-a_d) \cdot P_{sp}(l,k) \quad (26)$$

ただし，$P_{sp}(l,k)$ は音声存在確率であり，次式で表される．

$$P_{sp}(l,k) = \frac{|X(l,k)|^2}{P_{min}(l,k)} \quad (27)$$

ここで，$P_{min}(l,k)$ は雑音混入音声のパワースペクトルの極小値であり，以下のように求める．まず，雑音混入音声の平滑化パワースペクトル $P(l,k)$ を式 (28) のように求める．

$$P(k,l) = \eta P(l-1,k) + (1-\eta)|X(l,k)|^2 \quad (28)$$

ここで，η は平滑化定数である．次に，雑音混入音声のパワースペクトルの極小値をトラッキングする．このトラッキングは，窓の長さに関係なく，前のスペクトルの値の連続平均をとることでその極小値を求める [6],[11].

$$P_{min}(l,k) = \gamma \cdot P_{min}(l-1,k) + \frac{1-\gamma}{1-\beta}(P(l,k) - \beta \cdot P(l-1,k)) \quad (\text{If } P_{min}(l-1,k) \le P(l,k)) \quad (29)$$

$$P_{min}(l,k) = P(l,k) \quad (\text{If } P_{min}(l-1,k) > P(l,k)) \quad (30)$$

ここで，β と γ は実験的に決定した定数であり，また β は極小値の適応時間を制御する．

4 高速追従形雑音スペクトル推定法の改良

本節では，3 節で述べた方法において，VAD 法，及び，雑音スペクトル推定法を改良した方式について述べる．

4.1 VAD の改良

従来法では，式 (19) において正の定数 C を加えることで，雑音が大きい場合でも音声スペクトルが小さい場合でも，音声区間と無音区間のエントロピーの変化を大きくし，VAD の精度を高めていた [7]．しかし，この正の定数 C は入力 SNR_{seg} によって最適な値が存在し，従来法ではこの最適な値を求めることができない．そこで我々は，入力 SNR_{seg} に応じて最適な値を求めることで，VAD の精度を高め，より正確に雑音スペクトルを推定できるよう VAD を改良した．

C は次のように求めることができる．先頭数フレーム間の全ての周波数帯において，式 (20) によって表される $X_{energy}(l,k)$ の最大値を求め，その 2 乗に係数 k を掛けたものを新たな C とする．

$$C = k \cdot \{max(X_{energy}(l,k))\}^2 \quad 1 \le l \le 5 \quad (31)$$

4.2 雑音スペクトルの推定

4.2.1 無音フレーム

無音区間における提案法の趣旨は「従来法の音声区間における雑音スペクトル推定法に準じるが，式 (26) の平滑化パラメータは，非定常な雑音に対してはうまく雑音スペクトルを推定できないという問題があるので，提案法では，フレームごとの音声存在確率にともなって，連続的に変化する平滑化パラメータを導入する．以下に提案法における無音フレームにおける雑音スペクトルの推定法を示す．

まず，式 (27) から音声存在確率を決定したら，平滑化パラメータ $\delta(l,k)$ を計算する．この平滑化パラメータは，そのフレームの音声存在確率と連続的に変化するシグモイド関数として計算される．

$$\delta(l,k) = \frac{1}{1+\exp(-r \cdot (P_{sp}(l,k) - t \cdot T_p(l,k)))} \quad (32)$$

ここで t は定数，$T_p(l,k)$ はそのフレームの適応閾値である．$T_p(l,k)$ は音声フレームで次のように計算される．

$$T_p(l,k) = \frac{|X(l,k)|^2_{mean}}{\bar{N}_{mean}(l,k)} \quad (33)$$

$$|X(l,k)|^2_{mean} = E[|X(i,k)|^2] \quad (34)$$

$$\bar{N}_{mean}(l,k) = E[\bar{N}(i,k)] \quad (35)$$

(i ∈ all speech-present frames,up to l_{th}frame)

雑音スペクトルの推定値は，平滑化パラメータ $\delta(l,k)$ を用いて，

$$\bar{N}(l,k) = \delta(l,k) \cdot \bar{N}(l-1,k) + (1-\delta(l,k)) \cdot |X(l,k)|^2 \quad (36)$$

と計算する．雑音スペクトルは時間的に変化するので式 (24) では推定がうまくいかなかったが，式 (36) ではフレームごとに適応閾値を導入し，音声存在確率にともなって連続的に変化する平滑化パラメータを用いているので，従来法よりも正確に雑音スペクトルを推定できる．

4.2.2 音声フレーム

従来法では，Cohen のアルゴリズム [10] を用いて式 (25) のように平滑化パラメータを適応的に制御しながら雑音スペクトルを推定していたが，非定常な雑音に対しては推定がうまくいかなかった．そこで，音声フレームでは，従来法でも追従能力が高い重み付き雑音推定法 [3],[4],[8] に基づいて雑音スペクトルを推定する．

重み付き雑音推定では，事後 SNR$\gamma(l,k)$ の推定値に応じて重み付けした雑音混入音声を用いて，継続的に雑音推定値を更新する．このため，過大推定を防ぎつつ，非定常雑音に対して高い追従性を達成する．

重み付き雑音推定は，SNR の推定，重み係数の計算，平均化処理で構成される．まず最初に，事後 SNR$\gamma(l,k)$ の推定値を式 (8) によって求め，これをもとに，図 2 の非線形関数を用いて重み係数 $W(l,k)$ を計算する．この非線形関数は，重み付け要素が SNR 推定値に反比例するようにデザインされている．このために，高 SNR に対する過大推定が防止される．

推定雑音スペクトル $\bar{N}(l,k)$ は，重み付けされた雑音混入音声 $z(l,k)$ の平均値としてそれぞれ次式で表せる．

$$z(l,k) = W(l,k) \cdot |X(l,k)|^2 \quad (37)$$

$$\bar{N}(l,k) = \frac{trace\{\mathbf{Z}(l,k)\}}{\psi(\mathbf{Z}(l,k))} \quad (38)$$

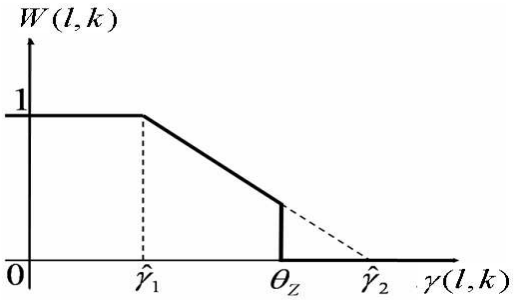

図 2: 重み係数 $W(l,k)$

$$\mathbf{Z}(l,k) = \begin{cases} \left[z(l,k), \tilde{\mathbf{Z}}(l-1,k)\right], & l < T_{init} \\ \left[z(l,k), \tilde{\mathbf{Z}}(l-1,k)\right], & \hat{\gamma}(l,k) < \theta_Z \\ \mathbf{Z}(l-1,k), & otherwise \end{cases} \quad (39)$$

$$\tilde{\mathbf{Z}}(0,k) = \mathbf{0}_{1\times(L_z-1)} \quad (40)$$

$$\tilde{\mathbf{Z}}(l,k) = \mathbf{Z}(l,k)\left[\mathbf{I}_{L_z-1}\mathbf{0}^T_{1\times(L_z-1)}\right]^T \quad (41)$$

式 (38) における $\psi(\mathbf{Z}(m,k))$ は $\mathbf{Z}(m,k)$ における非ゼロの要素数を表し，$trace\{\mathbf{Z}(m,k)\}$ は行ベクトル $\mathbf{Z}(m,k)$ の要素の総和となる．ただし，L_z は重み付けされた雑音混入音声 $z(m,k)$ を平均化する時のサイズであり，$\mathbf{I}_{L_z-1}$ は大きさ L_z-1 の単位行列であり $\mathbf{0}^T_{1\times(L_z-1)}$ は大きさ L_z-1 の零ベクトルである．また，最初の T_{init} フレームを無音区間と仮定し，雑音スペクトルの初期推定値を求める．

5 シミュレーション

入力信号として，8kHz で標本化された男性及び女性の音声を用いた．雑音としては 10000 サンプルまでは非定常な雑音であるバブル雑音を付加し，10001〜30000 サンプルでは定常な雑音である白色雑音を付加した．

5.1 評価方法

5.1.1 正規化推定誤差

雑音スペクトル推定精度の評価として，フレームごとに次の式 (42) で与えられる正規化推定誤差 $\varepsilon(l)$ を用いて評価した．

$$\varepsilon(l) = 10\log_{10}\left(\frac{\sum_{k=0}^{M}\left||D(l,k)|^2 - |\bar{N}(l,k)|^2\right|}{\sum_{k=0}^{M}|D(l,k)|^2}\right) \quad (42)$$

$$\bar{\varepsilon} = \frac{1}{L}\sum_{l=1}^{L}\varepsilon(l) \quad (43)$$

ただし，L は全フレーム数である．上式の ε は，値が小さいほど雑音スペクトル推定が正確であるということを表している．また，$\bar{\varepsilon}$ は全フレームの正規化推定誤差 $\varepsilon(l)$ の平均値を表している．

5.1.2 SNR 評価

出力では，信号を 12ms の区間に分割し，各区間の SNR の平均を求めるセグメンタル SNR で評価を行なう．SNR_{seg} は各区間の SNR の平均を求める評価法である．音声信号は時々刻々と変化しているので，細かい時間間隔で SNR を求め，その平均値である SNR_{seg} は，雑音が低エネルギーで広域に分布している場合，雑音除去性能を正しく評価を行なうことができる．セグメンタル SNR は次式で定義される．

$$SNR_{seg} = \frac{10}{L}\sum_{l=0}^{L-1}\log_{10}\frac{\sum_{n=N_l}^{N_l+N-1}s^2(n)}{\sum_{n=N_l}^{N_l+N-1}(\hat{s}(n)-s(n))^2} \quad (44)$$

ただし，N は分析フレームの長さである．

5.1.3 理想値

理想値として，正確な雑音スペクトルを用いてスペクトルゲインを計算し，雑音抑圧音声を求めた場合を理想値とする．正確な雑音スペクトルで計算したゲインを $G_{tl}(m,k)$ とすると，理想値の雑音抑圧音声は次式となる．

$$\hat{s}(n) = IFFT\left[G_{tl}(l,k)|X(l,k)|\exp\left(j\theta(l,k)\right)\right] \quad (45)$$

5.2 シミュレーション結果・考察

入力 $SNR_{seg} = 0$db のときの正規化推定誤差を図 3 に，また周波数が 3kHz のときの雑音スペクトルを図 4 に示す．

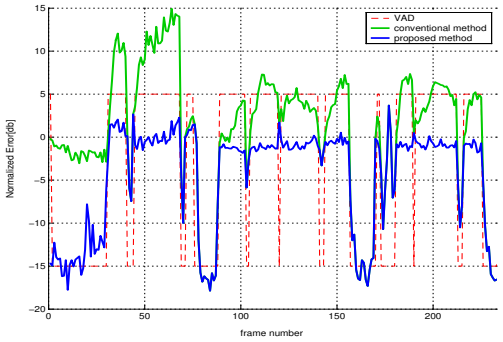

図 3: 正規化推定誤差 $\varepsilon(l)$

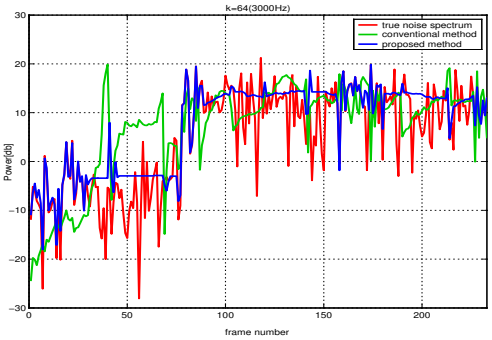

図 4: 雑音スペクトル (周波数 3kHz)

図 3 の赤点線は VAD によって検出した音声の有無を表していて，赤点線が立ち上がっている場合は音声フレーム，逆に下がっている場合は無音フレームを表している．図 3 を見れば分かるように，無音フレームにおいては音声フレームよりも雑音スペクトル推定が正確に行なえていることが分かる．図 4 の赤線は実際の雑音スペクトル，青線は推定した雑音スペクトルを表している．約 80 フレームで雑音スペクトルが急変化していることがわかるが，推定した雑音スペクトルはこの変化に追従しながら雑音スペクトルを推定できていることが分かる．

また，各入力 SNR_{seg} における出力 SNR_{seg} と正規化推定誤差平均を表 1 と表 2 に示す．これらの表より提案法は従来法に比べて，出力 SNR_{seg}・正規化推定誤差平均ともに向上していることが分かる．

表 1: 出力 SNR_{seg}[dB]

入力 SNR_{seg}[dB]	0	3	6	9
MMSE STSA(理想値)	10.58	12.34	14.32	16.45
MMSE STSA(従来法)	3.147	5.254	6.859	8.057
MMSE STSA(提案法)	5.314	7.475	9.642	11.89
Joint MAP(理想値)	10.58	12.36	14.33	16.42
Joint MAP(従来法)	3.180	5.612	7.450	8.904
Joint MAP(提案法)	5.388	7.595	9.972	12.30

表 2: 正規化推定誤差平均

入力 SNR_{seg}[dB]	0	3	6	9
従来法	-0.4373	-0.6667	0.4369	2.152
提案法	-4.517	-3.864	-3.187	-2.521

6 まとめ

本稿では，スペクトルサプレッション法における雑音スペクトル推定について検討を行った．特に，雑音スペクトルの急激な変化に追従できる方式を検討した．具体的には，無音区間と音声区間の検出方法の改善，各区間における雑音スペクトル推定に対する最適な方式の選択及び改善を行った．その結果，雑音が時間的に急激に変化しても，その変化に追従し，かつ良好に雑音スペクトルが推定できた．

今後の課題としては，音声区間における雑音スペクトルの推定精度の向上，及び種々の雑音における有効性の確認が挙げられる．

参考文献

[1] Y.Ephraim and D.Malah, “Speech enhancement using minimum mean-square error short-time spectral amplitude estimator”, IEEE Trans Acoust., Speech, Signal Processing, ASSP-32, 6, pp.1109-1121, Dec.1984.

[2] T.Lotter and P.Vary, “Noise reduction by joint maximum a posteriori spectral amplitude and phase estimation with super-gaussian speech modeling”, Proc. EUSIPCO-04(Vienna,Austria), pp.1447-60, Sep.2004.

[3] M.Katou, A.Sugiyama and M.Serizawa, “Noise suppression with high speech quality based on weighted noise estimation and MMSE STSA”, IEICE Trans.Fundamental, vol.E85-A, no.7, pp.1710-1718, Jul.2002.

[4] 鈴木大和, 中山謙二, 平野晃宏, “スペクトルサプレッション法における無音区間の検出と雑音スペクトル推定の改善”, 第 2 1 回信号処理シンポジウム（京都）, C3-2, 2006.11.

[5] R.Martin, D.Malah, V.Cox and J.Accardi , “A noise reduction preprocessor for mobile voice communication”, EURASIP Journal on Applied Signal Processing, pp.1046-1058, Aug.2004.

[6] B.F.Wu, K.C.Wang, and L.Y.Kuo, “A noise estimator with rapid adaptation in variable-level noisy environments”, Proceeding ROCLING XVI, Taipei, sep.2004.

[7] C.Jia and B.Xu, “An improved entropy-based endpoint detection algorithm”,
Proc. Int. Sympo. Chinese Spoken Language Processing, pp.1399-1402, Aug. 2002.

[8] 大和一洋, 杉山昭彦, 加藤正徳, “Post-processing noise suppresor with adaptive gain-flooring suitable for distorted speech”, 電子情報通信学会 2006 年ソサイエティ大会, 金沢, A-4-20, pp.87, Sep.2006.

[9] 鈴木大和, 中山謙二, 平野晃宏, “スペクトルサプレッション法によるノイズキャンセラの音質改善”, 信学技法，SIP2005-11, May. 2005.

[10] I.Cohen and B.Berdugo, “Noise estimation by minima controlled recursive averaging for robust speech enhancement”, IEEE Signal Process. Lett. 9(1), 12-15, 2002.

[11] G.Doblinger, “Computationally efficient speech enhancement by spectral minima tracking in subbands”, EUSPICO , pp.1513-1516, 1995.