

非定常雑音環境下における VAD を用いた高速雑音推定法

A Noise Spectral Estimation Method Based on VAD with Rapid Adaptation for Non-stationary Noise Environments

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 Science
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いて、雑音の種類や大きさが急激に変化したときにその変化に追従できる雑音スペクトル推定を提案する。提案法では、まず Voice Activity Detector（以下、VAD とする）において2つの閾値を設け、各フレームのスペクトルエントロピーを比較することで、無音区間・準音声区間・音声区間に分類する。その後それぞれの区間において、適切な雑音スペクトル推定法を用いて雑音スペクトルを推定する。シミュレーションでは、NOISEX-92 から入手した White, Babble, Car, Pink, Factory, Tank Noise の計6種類の雑音を使用し、雑音のレベルも時間的に変動させている。提案法を用いることで雑音の急変化にも対応でき、また従来法よりも正規化推定誤差と SNR が向上している。

ABSTRACT

A noise spectral estimation method, which is used in spectral suppression noise cancellers, is proposed for highly nonstationary noise environments. Speech, quasi-speech and non-speech frames are detected by using the entropy-based voice activity detector (VAD). An adaptive normalization parameter and a variable two thresholds are newly introduced for the VAD. They are very useful for rapid change in the noise spectrum and power. Furthermore, an adaptive method is applied to estimating the noise spectrum in each frame. Simulations are carried out by using many kinds of noises, including white, babble, car, pink, factory and tank, which are changed from one to the other. The segmental SNR is improved by 2.9~5dB, and noise spectral estimation error is improved by 9.8~11dB for the babble noise and the other noise, which are changed from one to the other.

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、雑音混入音声のスペクトルと雑音スペクトルの比を雑音混入音声に乗じて雑音成分を抑制するスペクトルサプレッション法が研究されている [1]-[4],[9]。

スペクトルサプレッション法において、雑音の抑圧度を定めるスペクトルゲインの計算方法として Joint MAP 法 [2] がよく用いられるが、この方法で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

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、

$$x(n) = s(n) + d(n) \quad (1)$$

と表せる。音声信号は一般に非定常でありその音響的特

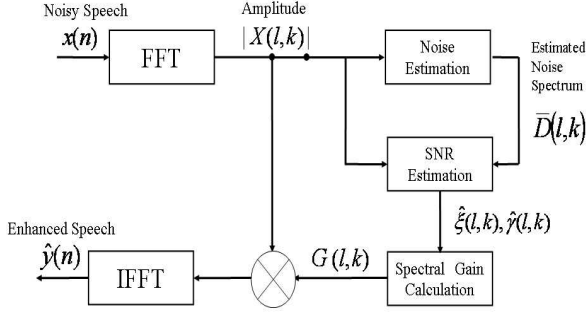


図 1: スペクトルサプレッション法のブロック図

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 番目のフレームに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．

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

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 は推定しなくてはならない．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．

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

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

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

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 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，

$$G(l, k) = u(l, k) + \sqrt{u^2(l, k) + \frac{\tau}{2\hat{\gamma}(l, k)}} \quad (9)$$

$$u(l, k) = \frac{1}{2} - \frac{\mu}{4\sqrt{\hat{\gamma}(l, k)\hat{\xi}(l, k)}} \quad (10)$$

と求められる．

3 従来の高速追従型雑音スペクトル推定法

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

3.1 VAD -Voice Activity Detector-

VAD 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．

$$H(l) = - \sum_{k=1}^{2M} P_r(l, k) \cdot \log(P_r(l, k)) \quad (11)$$

$$P_r(l, k) = \frac{|X(l, k)|^2}{\sum_{k=1}^{2M} |X(l, k)|^2} \quad (12)$$

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．

$$|X(l, k)|^2 = 0, \quad k \leq 250\text{Hz} \text{ or } k \geq 4000\text{Hz} \quad (13)$$

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 (12) 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 (12) において、正の定数 C を加えることで、雑音が多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]。

$$\hat{H}(l) = - \sum_{k=1}^{2M} \hat{P}_r(l, k) \cdot \log(\hat{P}_r(l, k)) \quad (14)$$

$$\hat{P}_r(l, k) = \frac{|X(l, k)|^2 + C}{\sum_{k=1}^{2M} |X(l, k)|^2 + C} \quad (15)$$

3.2 Rapid Adaptation

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。

$$\hat{D}(l, k) = \lambda \cdot \hat{D}(l-1, k) + (1-\lambda) \cdot |X(l, k)|^2 \quad (16)$$

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 (17) 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。

$$\hat{D}(l, k) = \rho(l, k) \cdot \hat{D}(l-1, k) + (1-\rho(l, k)) \cdot |X(l, k)|^2 \quad (17)$$

$$\rho(l, k) = a_d + (1-a_d) \cdot P_{sp}(l, k) \quad (18)$$

ただし、 $P_{sp}(l, k)$ は音声存在確率であり、次式で表される。

$$P_{sp}(l, k) = \frac{|X(l, k)|^2}{P_{min}(l, k)} \quad (19)$$

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 (20) のように求める。

$$P(k, l) = \eta P(l-1, k) + (1-\eta) |X(l, k)|^2 \quad (20)$$

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]。

$$P_{min}(l, k) = \gamma \cdot P_{min}(l-1, k) + \frac{1-\gamma}{1-\beta} (P(l, k) - \beta \cdot P(l-1, k)) \quad (\text{If } P_{min}(l-1, k) \leq P(l, k)) \quad (21)$$

$$P_{min}(l, k) = P(l, k) \quad (\text{If } P_{min}(l-1, k) > P(l, k)) \quad (22)$$

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 (12) において正の定数 C を加えることで、SNR が低い場合でもエントロピーの変化を大きくし、より正確に無音区間と音声区間を検出していた [7]。しかしこの正の定数 C は、入力 SNR_{seg} や雑音の種類によって適切な値が存在するので、従来法では雑音の種類や大きさが変化した時にこの適切な値を求めることができない。また、文献 [12] の C_{new} では、バブル雑音以外の雑音では、適切な値を求めることができなかった。

さらに従来法の VAD では、閾値付近に存在するエントロピーの情報だけでは、本当に無音区間か音声区間かは定かではない。

そこで我々は、まず正の定数 C について、入力 SNR_{seg} に応じて適切な値が随時更新される新たな C を以下の式で表される C_{new} として導入する。

$$C_{new}(l) = MAX\{|X(l, k)|\}, 250 \leq k \leq 4000 \quad (23)$$

以上のように C_{new} をフレーム毎に更新することで、雑音の種類や大きさが変化した場合でも適切なパラメータが計算されるので、VAD の精度が高まり正確に雑音スペクトルを推定できるようになる。

また提案法の VAD では閾値を 2 つ設け、かつこれらの 2 つの閾値を各フレーム毎に更新する。両閾値はまず、先頭 5 フレームのエントロピーの平均値にそれぞれ異なる定数 c_1, c_2 (但し $c_1 > c_2$) を掛けたものを最初の閾値 $\sigma_1(1), \sigma_2(1)$ とする。そしてこれらの閾値は、現フレームのエントロピーが一つ前のフレームの閾値より大きかった場合、前のフレームで計算した 5 フレーム分のエントロ

ピーの中で最も過去にある値を破棄し、現フレームのエントロピーと過去4フレーム分のエントロピーの平均値に定数 c_1, c_2 を掛けたものを現フレームの閾値 $\sigma_1(l), \sigma_2(l)$ として更新する。

なお、現フレームのエントロピーが閾値 $\sigma_1(l)$ より大きければ無音区間、閾値 $\sigma_2(l)$ より大きい閾値 $\sigma_1(l)$ より小さい場合は半無音区間、閾値 $\sigma_2(l)$ より小さい場合は音声区間と判断する。

提案法では閾値を2つ導入して、今まで無音区間か音声区間か曖昧だった区間を半無音区間とし、3つの区間でそれぞれ適切な雑音推定アルゴリズムを用いることで、より正確に雑音スペクトルを推定することができる。またこれら2つの閾値を各フレーム毎に計算することで、入力 SNR が小さい場合でも柔軟な区間判別が可能となり、また雑音の種類や大きさが変化した場合でも、それらの変化に追従して適切な閾値を設けることができる。

4.2 雑音スペクトルの推定

4.2.1 無音フレーム

式 (18) 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った。しかし提案法では、VAD で2つの閾値を用いることで、現フレームが無音フレームという信頼がおける場合のみ、そのフレームを無音区間と判断する。

完全な無音フレームにおける雑音混入音声スペクトルは、雑音スペクトルに等しいので、提案法における推定雑音スペクトルは以下のように定める。

$$\hat{D}(l, k) = |X(l, k)|^2 \quad (24)$$

4.2.2 準音声フレーム

我々が言及している半無音フレームは、無音フレームか音声フレームか曖昧であるようなフレームのことを指す。そこで半無音フレームでは式 (19) で計算できる音声存在確率を用いて平滑化パラメータを計算し、この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を行なう。

なお、雑音推定スペクトルは、前述の3.3.2節で述べた式 (17)～式 (22) で計算する。

4.2.3 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 (17) 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。

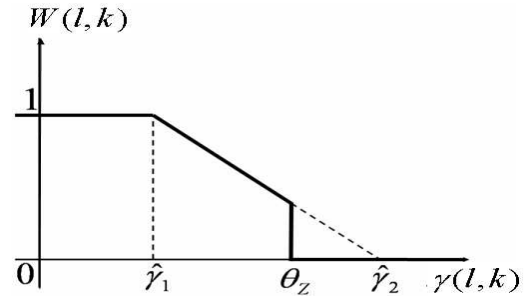


図2: 重み係数 $W(l, k)$

推定雑音スペクトル $\hat{D}(l, k)$ 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。

$$z(l, k) = W(l, k) \cdot |X(l, k)|^2 \quad (25)$$

$$\hat{D}(l, k) = \frac{\text{trace}\{\mathbf{Z}(l, k)\}}{\psi(\mathbf{Z}(l, k))} \quad (26)$$

式 (26) における $\psi(\mathbf{Z}(l, k))$ は $\mathbf{Z}(l, k)$ における非ゼロの要素数を表し、 $\text{trace}\{\mathbf{Z}(l, k)\}$ は行ベクトル $\mathbf{Z}(l, k)$ の要素の総和となる。

5 シミュレーション条件

入力信号として、8kHz で標準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 Babble Noise を付加し、10001～30000 サンプルでは前半の Babble Noise より強さが大きい White, Babble, Car, Pink, Factory, Tank Noise を付加した。

5.1 評価方法

5.1.1 雑音スペクトルの正規化推定誤差

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 (27) で与えられる正規化推定誤差 $\varepsilon(l)$ を用いて評価した [3]。

$$\varepsilon(l) = 10 \log_{10} \left(\frac{\sum_{k=0}^M \left| |D(l, k)|^2 - |\hat{D}(l, k)|^2 \right|}{\sum_{k=0}^M |D(l, k)|^2} \right) \quad (27)$$

$$\bar{\varepsilon} = \frac{1}{L} \sum_{l=1}^L \varepsilon(l) \quad (28)$$

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示す。

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 時間領域における音声推定誤差の評価

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ため、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。

$$SNR_{seg} = \frac{10}{L} \sum_{l=0}^{L-1} \log_{10} \frac{\sum_{n=N_l}^{N_l+N-1} s^2(n)}{\sum_{n=N_l}^{N_l+N-1} (\hat{s}(n) - s(n))^2} \quad (29)$$

ただし、 N は分析フレームの長さである。

5.1.3 音声スペクトル歪みの評価

音質の評価方法として、Log-Spectral Distortion を行なう。LSD は周波数領域においてクリーン音声の振幅値 $|S(l, k)|$ と、雑音抑圧音声の振幅値 $|\hat{S}(l, k)|$ の各分析フレームにおける比率の平均値を求めていて、次式で表される [9]。

$$LSD = \frac{1}{L} \sum_{l=1}^L \left(\frac{1}{2M} \sum_{k=1}^{2M} \left(\log \frac{|S(l, k)| + \delta}{|\hat{S}(l, k)| + \delta} \right)^2 \right)^{\frac{1}{2}} \quad (30)$$

δ は微小値、 $2M$ は分析フレーム長である。LSD は常に正の値をとり、また値が小さいほど音質が良いことを表している。人間の聴覚に対して位相情報はあまり影響を与えないので、音質の評価方法としては位相情報も含まれる SNR_{seg} よりも、式 (30) で示される LSD の方が有効な評価方法と言える。

5.1.4 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 $G_{tl}(l, k)$ とすると、理想値の雑音抑圧音声は次式となる。

$$\hat{s}(n) = IFFT[G_{tl}(l, k)|X(l, k)| \exp(j\theta(l, k))] \quad (31)$$

5.2 シミュレーション結果・考察

付加雑音として Babble Noise(6dB) と White Noise(2dB) を付加し、全体の入力 SNR_{seg} が 3dB のときの正規化推定誤差を図 3 に、また周波数が 2kHz のときの雑音スペクトルを図 4 に示す。

図の点線は VAD によって検出した音声の有無を表している。点線が立ち上がっている場合は音声区間、逆に下

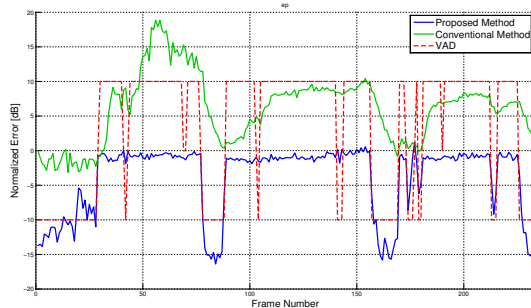


図 3: 正規化推定誤差 $\varepsilon(l)$ の時間的変化

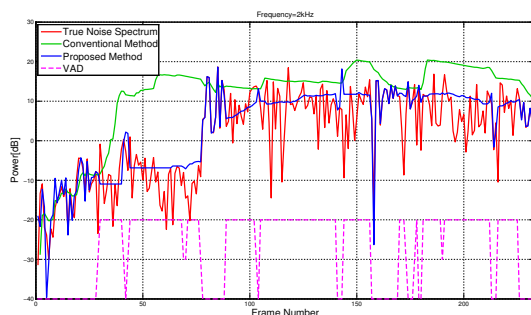


図 4: 雑音スペクトルの時間的変化 (周波数 2kHz)

がっている場合は無音区間、また点線の最大値と最小値の真ん中に位置する点線は半無音区間を表している。図 3 を見れば分かるように、無音区間においては音声区間よりも雑音スペクトル推定が正確に行えていることが分かる。また、音声区間においてもわずかに提案法の方が誤差が小さくなっている。

図 4 を見ると約 80 フレームで雑音スペクトルが急変化していることが分かる。従来法で推定した緑線の雑音スペクトルは、非定常なバブル雑音の推定がうまくいっていないが、提案法で推定した青線の雑音スペクトルは、雑音の種類によらず良好な推定が行なえており、また雑音の急変化に追従しながら推定できていることが分かる。

また、バブル雑音 (6dB)+他雑音 (2dB) を付加し、全体を 3dB としたときの正規化推定誤差平均、出力 SNR_{seg} 、LSD を、表 1 に理想値、表 2 に従来法、表 3 に提案法として示す。これらの表より提案法は従来法に比べて、正規化推定誤差平均・出力 SNR_{seg} ・LSD とともに向上していることが分かる。

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。具体的には、VAD の改良による無音区間・準音声区間・音声区間の検出方法の導入、各区間における雑音スペクトル推定に対する最

表 1: 音声品質の評価と正規化推定誤差 $\bar{\varepsilon}$ (理想値)

Evaluation Method	$\bar{\varepsilon}$	$SNR_{seg}[dB]$	LSD
Babble+White	-	10.04	0
Babble+Babble	-	11.28	0
Babble+Car	-	17.57	0
Babble+Pink	-	11.19	0
Babble+Factory	-	11.42	0
Babble+Tank	-	14.74	0

表 2: 音声品質の評価と正規化推定誤差 $\bar{\varepsilon}$ (従来法)

Evaluation Method	$\bar{\varepsilon}$	$SNR_{seg}[dB]$	LSD
Babble+White	6.097	4.592	1.141
Babble+Babble	8.679	3.783	1.110
Babble+Car	6.596	5.236	1.095
Babble+Pink	6.818	4.807	1.026
Babble+Factory	7.429	4.389	1.059
Babble+Tank	6.779	4.921	1.058

表 3: 音声品質の評価と正規化推定誤差 $\bar{\varepsilon}$ (提案法)

Evaluation Method	$\bar{\varepsilon}$	$SNR_{seg}[dB]$	LSD
Babble+White	-3.674	7.527	0.9578
Babble+Babble	-2.262	7.225	0.8468
Babble+Car	-3.578	10.79	0.7693
Babble+Pink	-3.631	8.322	0.8008
Babble+Factory	-3.224	8.143	0.8100
Babble+Tank	-3.414	9.904	0.7788

適な方式の選択及び改善である。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。

しかし、音声区間における雑音スペクトルの推定精度が無音区間に比べると劣っているので、今後の課題として音声区間における雑音スペクトルの推定精度の向上が上げられる。

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", IE-ICE 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 21 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 end-point 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 suppressor 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おける雑音スペクトル推定とスペクトルゲイン制御に関する研究", 金沢大学大学院自然科学研究科電子情報工学専攻 修士論文, May. 2007.

[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 sub-bands", EUSPICO, pp.1513-1516, 1995.

[12] 東尚哉, 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-", 第 22 回信号処理シンポジウム (仙台), B4-2, 2007.11.