

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

Improvement of Spectrum Suppression Noise Cancelers Based on Detection
of Non-Speech Section and Noise Spectrum Estimation

鈴木 大和[†] 中山 謙二[‡] 平野 晃宏[‡]
Hirokazu Suzuki Kenji Nakayama Akihiro Hirano

[†]金沢大学大学院 自然科学研究科 電子情報工学専攻

Division of Electronics and Computer Science

Graduate School of Natural Science and Technology, Kanazawa Univ.

[‡]金沢大学大学院 自然科学研究科 電子情報科学専攻

Division of Electrical Engineering and Computer Science

Graduate School of Natural Science and Technology, Kanazawa Univ.

E-mail: 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ノイズキャンセラにおける雑音スペクトル推定の改善について検討する。音声が存在する区間でも雑音推定が可能である重み付き雑音推定において推定された雑音スペクトルと実際の雑音スペクトルとでは誤差が生じる。特に、無音区間で誤差が大きい場合、雑音が大きく残ってしまう。そこで、周波数領域において入力信号における各フレームのエントロピーの値を用いて、無音区間をより正確に検出することで雑音スペクトル推定、SNR推定、及びスペクトルゲインの補正を行う方法を提案する。この方法により、背景雑音が定常な雑音である白色雑音、非定常な雑音であるパブル雑音の両方の場合において、従来の方法よりもSNRが向上した。

ABSTRACT

A single microphone noise canceller based on a spectrum suppression approach is taken into account. A detection method of unvoiced sections is introduced and an improved method for noise spectrum estimation is proposed. The noise spectrum can be more precisely estimated compared with the conventional methods. A spectrum gain is also improved more precisely by using the noise spectrum estimated by the proposed method. It is confirmed that the segmental SNR can be improved by 1~1.5 dB for stationary white noise and non-stationary bubble noise through computer simulations.

1 まえがき

現在、携帯電話などの移動通信が広く普及し、ほとんどの人が携帯電話を持つようになっている。そのため、街頭や車内など背景雑音が多い場所において携帯電話が使用される機会が増えている。このような雑音が利用者の会話中にマイクに拾われると、相手側では会話が聞きづらくなり快適な使用の妨げとなる。そこで、雑音を除去するためにノイズキャンセラが必要である。その方式には、単一マイクを用いる方式、つまり、音声+雑音を集音するマイクのみを用いる方式と、複数のマイクを用いる方式、つまり、音声+雑音を集音するマイク以外に主に雑音のみを集音するマイクを用いる方式がある。携帯電話など、小型・低電力が要求される場合には単一マイク方式のノイズキャンセラが望ましい。

様々なノイズキャンセラの方式が提案されているが、ETSI(European Telecommunications Standard Institute)が定める3GPP(The 3rd Generation Partnership Project)における携帯電話用雑音抑圧技術の標準化においては、ノイズキャンセラの方式は推奨されておらず、その性能に関する要求条件とその評価方法のみが定められている [3],[4]。

本稿では、携帯電話への応用を考えて、単一マイク方式のノイズキャンセラを対象とし、スペクトルサプレッション法を採用する。雑音の抑圧度を定めるスペクトルゲインを求める方法として、Ephraim と Malah によって提案された MMSE STSA 法がある [1]。MMSE STSA 法は、統計モデルとして、音声スペクトルをガウス分布であると仮定してスペクトルゲインを求める。し

しかし、音声スペクトルの確率分布はガウス分布とは異なっている。そこで、より実際の音声スペクトルの分布に近いスーパーガウス分布を仮定して、スペクトルゲインを求める Joint MAP 法が Lotter と Vary によって提案されており、MMSE STSA 法よりも良好な結果を示している [5]。

これらのスペクトルサブ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、雑音スペクトルが必要である。雑音スペクトルの推定が不正確であると、抑圧後に雑音が大きく残ってしまったり、雑音の過大推定により、雑音抑圧後に音声が大きくひずむことになり、音質の低下を招く原因となる。このようなことから、雑音スペクトルをいかに正確に求められるかがスペクトルサブプレッション法において重要である。雑音スペクトル推定として、雑音の特性変動に追従して雑音スペクトルを推定できる重み付き雑音推定がある [2]。この方式は 3GPP が定める次世代の携帯電話雑音除去性能を満たしており、市販の製品にも搭載されている [4]。この方式では、推定された雑音スペクトルと実際の雑音スペクトルで誤差が生じる。特に、雑音の特性変動が大きいバブル雑音では誤差が大きく、無音区間で誤差が大きい場合、雑音が大きく残ってしまう。そこで、本稿では周波数領域において、入力信号（雑音混入音声）における各フレームのエントロピーの値を用いて、無音区間をより正確に検出する方法 [7]～[10] を採用し、雑音スペクトル推定、SNR 推定、及びスペクトルゲインの補正をより正確に行なう方法を提案する。

2 スペクトルサブプレッション法

2.1 スペクトルサブプレッション法の構成

図 1 にスペクトルサブプレッション法のブロック図を示す。

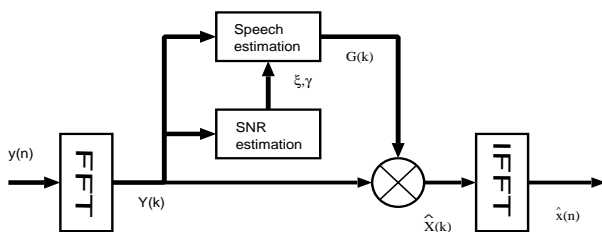


図 1: スペクトルサブプレッション法のブロック図

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あるとする。時間領域でのクリアな音声 $x(n)$ と雑音 $d(n)$ より雑音混入音声 $y(n)$ は次式で表せる。

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

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。そのために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下で、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 演算を実行して、スペクトルの時系列を得ている。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、 $2M$ 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、 m 番目のフレームにおける切り出された雑音混入音声 $\hat{y}_m(n)$ は、

$$\hat{y}_m(n) = \begin{cases} h(n)y_{m-1}(n) & 1 \leq n \leq M \\ h(n)y_m(n-M) & M \leq n \leq 2M \end{cases} \quad (2)$$

と表せる。

m 番目のフレームにおける k 番目の周波数成分は次式で表せる。

$$Y(m, k) = X(m, k) + D(m, k) \quad (3)$$

$$Y(m, k) = R(m, k) \exp(j\theta(m, k)) \quad (4)$$

$$X(m, k) = A(m, k) \exp(j\alpha(m, k)) \quad (5)$$

雑音スペクトルを $\lambda_d(m, k)$ とすると、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) はそれぞれ次式で表せる。

$$\xi(m, k) = \frac{A^2(m, k)}{\lambda_d(m, k)} \quad (6)$$

$$\gamma(m, k) = \frac{R^2(m, k)}{\lambda_d(m, k)} \quad (7)$$

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 $\xi(m, k)$ は、次式で推定できる [1]。

$$\hat{\xi}(m, k) = \alpha_{SNR} \gamma(m-1, k) G^2(m-1, k) + (1 - \alpha_{SNR}) P[\gamma(m, k) - 1] \quad (8)$$

ただし、 α_{SNR} は $0 < \alpha_{SNR} < 1$ を満たす係数であり、 $P[x]$ は次式で表せる。

$$P[x] = \begin{cases} x, & x > 0 \\ 0, & otherwise \end{cases} \quad (9)$$

また、雑音スペクトルは、文献 [2] によって提案された重み付き雑音推定によって求める。推定された雑音を下に、事後 SNR $\gamma(m, k)$ の推定値を求める。このようにして推定された事前 SNR、事後 SNR によりスペクトルゲイン $G(m, k)$ を求め、それを雑音混入音声に乗じることにより雑音を抑える。

2.2 MMSE STSA 法

MMSE STSA 法 [1] は雑音を含む音声から明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。統計モデルとして、音声と雑音ともにス

ベクトル成分において統計的独立でガウス分布であると仮定する．スペクトルゲインは

$$G(m, k) = \left[(1 + \nu(m, k)) I_0\left(\frac{\nu(m, k)}{2}\right) + \nu(m, k) I_1\left(\frac{\nu(m, k)}{2}\right) \right] \cdot \frac{\Lambda(m, k) \sqrt{\pi \nu(m, k)}}{1 + \Lambda(m, k) 2\gamma(m, k)} \exp\left(-\frac{\nu(m, k)}{2}\right) \quad (10)$$

で求められる．式中の各関数は

$$\eta(m, k) = \frac{\xi(m, k)}{1 - q(m, k)} \quad (11)$$

$$\nu(m, k) = \frac{\eta(m, k)}{1 + \eta(m, k)} \gamma(m, k) \quad (12)$$

$$\Lambda(m, k) = \frac{1 - q(m, k)}{q(m, k)} \cdot \frac{\exp(\nu(m, k))}{1 + \eta(m, k)} \quad (13)$$

で求められる． $I_0(\cdot), I_1(\cdot)$ はそれぞれ 0 次と 1 次の修正ベッセル関数， $q(m, k)$ を事前確率といい，非音声成分，及びパワーが十分小さい音声成分が，雑音混入音声に含まれる確率を表す．

2.3 Joint MAP 法

Joint MAP 法 [5] は，雑音をガウス分布信号，音声をスーパーガウス分布信号であるという仮定の下でスペクトルゲインを計算する方法である．

スーパーガウス分布の確率密度関数 (PDF) は音声の分散を λ_s として式 (14) で表せる．

$$p(A(m, k)) = \frac{\mu^{\tau+1}}{\Gamma(\tau+1)} \frac{A^\tau(m, k)}{\lambda_s^{\tau+1}} \exp\left\{-\mu \frac{A(m, k)}{\lambda_s}\right\} \quad (14)$$

ただし，式中の $\Gamma(\cdot)$ はガンマ関数を表し， τ と μ の 2 つのパラメータにより， $p(A(m, k))$ の形状を決定する．

Joint MAP 法では，条件付確率 $p(A, \alpha|Y)$ を最大にする振幅 A と位相 α を求め，その値を推定値として利用する．スペクトルゲインは次式で表せる．

$$G(m, k) = u(m, k) + \sqrt{u^2(m, k) + \frac{\tau}{2\gamma(m, k)}} \quad (15)$$

$$u(m, k) = \frac{1}{2} - \frac{\mu}{4\sqrt{\gamma(m, k)\xi(m, k)}} \quad (16)$$

2.4 事前確率推定

非音声成分，及びパワーが十分小さい音声成分が，雑音混入音声に含まれる確率である事前確率 $q(m, k)$ は，式 (10) ~ (13) から，その値がスペクトルゲインの計算に大きな影響を与えると言える．そこで，事後 SNR $\gamma(m, k)$ に閾値 γ_{th} を設けて，周波数ごとに適切な事前確率を推定する方法が提案されている [6]．次式で事前確率を推定する．

$$q(m, k) = \alpha_q q(m-1, k) + (1 - \alpha_q) I(m, k) \quad (17)$$

ただし， $\gamma(m, k)$ が γ_{th} よりも大きい時は，パワーが十分大きい音声成分が雑音混入音声に存在するとして， $I(m, k) = 1$ とし， $\gamma(m, k)$ が γ_{th} よりも小さい時は，パワーが十分大きい音声成分が雑音混入音声に存在しないとして， $I(m, k) = 0$ とする．

2.5 重み付き雑音推定

重み付き雑音推定 [2] では，事後 SNR $\gamma(m, k)$ 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 $\gamma(m, k)$ の推定値は，雑音混入音声スペクトル $R^2(m, k)$ と 1 フレーム前の推定雑音スペクトル $\lambda_d(m-1, k)$ より

$$\hat{\gamma}(m, k) = \frac{R^2(m, k)}{\lambda_d(m-1, k)} \quad (18)$$

と求められる．次に，推定した SNR $\hat{\gamma}(m, k)$ を下に，重み係数 $W(m, k)$ を計算する．図 2 の非線形な関数が $W(m, k)$ の計算に用いられる．この非線形な関数は，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．

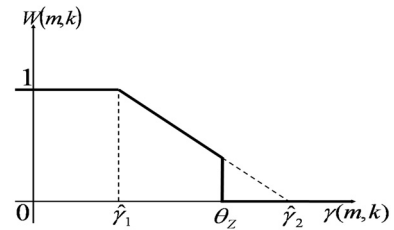


図 2: 重み係数 $W(m, k)$

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

$$z(m, k) = W(m, k) R^2(m, k) \quad (19)$$

$$\lambda_d(m, k) = \frac{\text{trace}\{\mathbf{Z}(m, k)\}}{\psi(\mathbf{Z}(m, k))} \quad (20)$$

ただし，

$$\mathbf{Z}(m, k) = \begin{cases} [z(m, k), \tilde{\mathbf{Z}}(m-1, k)] & , m < T_{init} \\ [z(m, k), \tilde{\mathbf{Z}}(m-1, k)] & , \hat{\gamma}(m, k) < \theta_Z \\ \mathbf{Z}(m-1, k) & , otherwise \end{cases} \quad (21)$$

$$\tilde{\mathbf{Z}}(0, k) = \mathbf{0}_{1 \times (L_z - 1)} \quad (22)$$

$$\tilde{\mathbf{Z}}(m, k) = \mathbf{Z}(m, k) [\mathbf{I}_{L_z - 1} \mathbf{0}_{1 \times (L_z - 1)}^T]^T \quad (23)$$

式(20)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, $\text{trace}\{\mathbf{Z}(m, k)\}$ 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 $\mathbf{0}_{1 \times (L_z - 1)}^T$ 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.

3 提案方式

本章では, 本稿で提案する雑音スペクトル推定の改善法について説明する. 重み付き雑音推定では, 1フレーム前の, 雑音スペクトルの推定値を下に決定された, 事後SNR $\gamma(m, k)$ の推定値に応じて, 重み係数が決定される. また, 雑音スペクトル推定において, 基準となる雑音スペクトルは, 入力信号の最初の区間を無音声区間と仮定し推定する. そのため, 推定値と実際の雑音スペクトルとは誤差が生じ, 特に, 無音区間で誤差が大きい場合, 雑音が大きく残る. そこで, 無音区間をより正確に検出することで, 雑音スペクトル推定, SNR推定, 及びスペクトルゲインの補正を行なう方法を提案する. まず, 無音区間の検出方法 [6] ~ [10] について述べ, 次に本稿で提案する方式について述べる.

3.1 VAD –Voice Activity Detector–

無音区間の検出方法として, 各フレームに対する全周波数における事後SNR $\gamma(m, k)$ の平均値 $\bar{\gamma}(m)$ を用いる方法が提案されている [6]. 定常で独立な雑音に対して, 無音区間であるフレームにおいて, 平均値 $\bar{\gamma}(m)$ はおよそ1となる. そこで, 各フレームにおいて平均値 $\bar{\gamma}(m)$ に対して, 適切な閾値を設定することにより無音区間を検出することができる. しかし, パルス雑音のような非定常な雑音の場合, 正確な無音区間の検出ができない.

そこで, 本稿では, 非定常な雑音でも無音区間の検出ができる, 入力信号のスペクトルエントロピー $H(m)$ を用い, 無音区間の検出を行なう方法 [7] ~ [10] を採用する. 無音区間では, スペクトルエントロピーは, 有音区間と比べて大きくなる. そこで, 入力信号の最初の区間を無音区間と仮定して, 最初のフレームから数フレーム分のスペクトルエントロピーの平均値を閾値とし, 閾値よりも小さい場合, 有音区間とする.

スペクトルエントロピーは次式で求める.

$$Y_{energy}(m, k) = R^2(m, k) \quad (24)$$

$$p(m, k) = \frac{Y_{energy}(m, k)}{\sum_{k=1}^{2M} Y_{energy}(m, k)} \quad (25)$$

$$H(m) = - \sum_{k=1}^{2M} p(m, k) \log(p(m, k)) \quad (26)$$

ただし, 式中の $2M$ は, 周波数データ数である. また, 音声スペクトルが, 最も集中している周波数帯域である $250Hz$ 以下, $4000Hz$ 以上では, $Y_{energy}(m, k) = 0$ とする. 論文 [9] より, スペクトルエントロピーの性質について以下のように報告されている.

- 式(25)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, 有音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, 有音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, 雑音が大きいか場合や, 音声スペクトルが小さい場合, 有音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区間の検出が難しくなる. しかし, 式(25)において, 次式のように正の定数 C を加えることにより, 雑音が大きいか場合でも, また, 音声スペクトルが小さい場合でも, 有音区間と無音区間でのエントロピーの変化を大きくすることができる [9].

$$\hat{p}(m, k) = \frac{(Y_{energy}(m, k) + C)}{\sum_{k=1}^{2M} (Y_{energy}(m, k) + C)} \quad (27)$$

$$\hat{H}(m) = - \sum_{k=1}^{2M} \hat{p}(m, k) \log(\hat{p}(m, k)) \quad (28)$$

3.2 雑音スペクトル推定の改善

無音区間のフレームにおいて, 重み付き雑音推定における重み係数を, 図3のような, 従来の重み係数よりも, 雑音混入音声の影響を強く受けるように設計された重み係数 $W_{new}(m, k)$ で, 再度雑音スペクトル推定及びSNR推定を行なう. ただし, 新しい重み係数での, 雑音推定の補正は, 長時間SNRが十分に高い場合は, 行なわないこととする. 長時間SNRを求めるために, ま

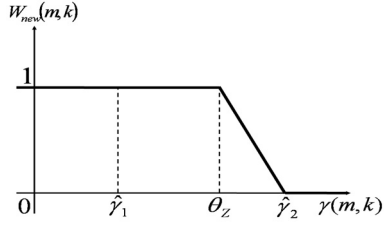


図 3: 新しい重み係数 $W_{new}(m, k)$

ず次式で信号のパワーの平均値を求める．

$$\bar{\lambda}_y(m) = \alpha_{LT} \bar{\lambda}_y(m-1) + (1 - \alpha_{LT}) \frac{1}{2M} \sum_{k=1}^{k=2M} R^2(m, k) \quad (29)$$

ただし, α_{LT} は, 忘却係数である．また, 式 (29) は, 有音区間でのみ計算を行なう． $\bar{\lambda}_y(m)$ と雑音スペクトル $\lambda_d(m, k)$ より, 長時間 SNR を求める [6]．

$$SNR_{LT}(m) = \frac{2M \bar{\lambda}_y(m)}{\sum_{k=1}^{2M} \lambda_d(m, k)} - 1 \quad (30)$$

3.3 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．スペクトルゲイン $G(m, k)$ に, 次に示す倍率 G_{sup} を乗じる．

$$G(m, k) = \begin{cases} G_{sup} G(m, k), & (\text{無音区間}) \\ G(m, k), & (\text{有音区間}) \end{cases} \quad (31)$$

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 [2],[14] 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すること [11],[12] を行なった．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．

$$G(m, k) = \begin{cases} G(m, k), & G(m, k) > G_{floor} \\ G_{floor}, & G(m, k) \leq G_{floor} \end{cases} \quad (32)$$

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 [14]．

4 シミュレーション

4.1 シミュレーション条件

入力信号としてサンプリング周波数 10kHz の男性音声を用いた．シミュレーションに使用する雑音として, 定常な雑音として白色雑音, 非定常な雑音としてパブル雑音 (公共の場所における非定常な雑音) とする．スペクトルゲイン及び, 重み付き雑音推定, SNR 推定に用いたパラメータを表 1 にまとめておく．

表 1: パラメーター

パラメーター	値	パラメーター	値
M	128	α_{SNR}	0.99
μ	1.74	τ	0.126
α_q	0.95	L_z	20
γ_1	0 dB	γ_2	10 dB
θ_z	7 dB	T_{init}	2
G_{sup}	0.25	α_{LT}	0.97
G_{floor} (有音区間)	-6.8 dB	G_{floor} (無音区間)	-10 dB

4.2 理想値の設定

スペクトルサプレッション法の理想値として, 雑音抑圧後のスペクトル $|\hat{X}(m, k)|$ として, クリーン音声のスペクトル $|X(m, k)|$ を用いた場合を理想値 1 とする．すなわち, 理想値 1 の雑音抑圧音声は次式となる．

$$\hat{x}(n) = IFFT[|X(m, k)| \exp(j\theta(m, k))] \quad (33)$$

次に, 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 2 とする．正確な雑音スペクトルで計算したゲインを $G_{ll}(m, k)$ とすると, 理想値 2 の雑音抑圧音声は次式となる．

$$\hat{x}(n) = IFFT[G_{ll}(m, k)|Y(m, k)| \exp(j\theta(m, k))] \quad (34)$$

4.3 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．セグメンタル SNR は次式で定義される．

$$SNR_{seg} = \frac{10}{L} \sum_{l=0}^{L-1} \log_{10} \frac{\sum_{n=N_l}^{N_l+N-1} x^2(n)}{\sum_{n=N_l}^{N_l+N-1} (\hat{x}(n) - x(n))^2} \quad (35)$$

ただし, N は分析フレームの長さ, L は全フレーム数である．また, $x(n)$ はクリーン音声である．

4.4 シミュレーション結果とその考察

表 2, 表 3 に白色雑音, パブル雑音の場合の出力におけるセグメンタル SNR を示す．

白色雑音, パブル雑音ともに従来法よりも SNR が向上している．特に, 入力 SNR が高くなるに従って, 出力での SNR が大きく向上している．このことは, 提案法により無音区間を検出し, 雑音スペクトル推定, 及びスペクトルゲインの補正を行なうことで, 有音区間での過剰抑圧を防ぎ, また, 無音区間で雑音を従来法よりも適切に抑えているためだと思われる．白色雑音と比較して雑音スペクトルの推定が難しいパブル雑音において従来法より SNR が大きく向上している．

表 2: 白色雑音に対する雑音抑圧後 SNR_{seg} [dB]

入力 SNR	6	12	18
理想値 1	14.15	18.59	23.37
MMSE(理想値 2)	11.71	15.42	19.64
MMSE(従来法 [1],[2])	10.63	14.05	17.46
MMSE(提案法)	11.24	14.95	18.65
Joint MAP(理想値 2)	11.86	15.80	20.32
Joint MAP(従来法 [2],[5])	10.71	14.46	18.35
Joint MAP(提案法)	11.39	15.16	18.97

表 3: バブル雑音に対する雑音抑圧後 SNR_{seg} [dB]

入力 SNR	6	12	18
理想値 1	12.91	17.63	22.57
MMSE(理想値 2)	10.04	14.09	18.89
MMSE(従来法 [1],[2])	8.07	12.43	16.78
MMSE(提案法)	8.93	13.58	17.85
Joint MAP(理想値 2)	10.21	14.58	19.53
Joint MAP(従来法 [2],[5])	8.22	12.99	17.75
Joint MAP(提案法)	9.06	13.79	18.18

5 まとめ

本稿では、スペクトルサブプレッション法による単一マイクノイズキャンセラにおいて、スペクトルエントロピーを用いて無音区間をより正確に検出することで、無音区間において以下の二つの改良を行なった。まず、従来の重み付き雑音推定 [2] における重み係数よりも、雑音混入音声の影響を強く受ける重み係数に変更し、雑音スペクトル推定、及び SNR 推定の補正を行なった。次に、スペクトルゲインを更に抑圧し、雑音をより抑えた雑音抑圧音声となるようスペクトルゲインの補正を行なった。さらに、無音区間、有音区間で適切にスペクトルゲインの最小値を設定すること [14]、及び原音付加を行なうこと [11],[12] により、雑音の過剰抑圧を防止した。シミュレーションで、従来法と提案法の比較を行なった結果、白色雑音、バブル雑音の両方において SNR が向上した。

今後の課題として、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理想値 2 に比べて、本稿で提案した方法は、SNR、音質ともに劣っているので、雑音スペクトル推定の精度の向上が望まれる。

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] 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.
- [3] 加藤正徳, 芹沢昌宏, 杉山昭彦, "携帯電話用雑音抑圧技術の標準化", 第 17 回 DSP シンポジウム, A5-1, Nov. 2002.
- [4] 加藤正徳, 芹沢昌宏, 杉山昭彦, 長谷川厚志, 水越明哉, "重み付き雑音推定に基づくノイズサブプレッサの DSP 実現と 3GPP 評価", 第 17 回 DSP シンポジウム, A5-3, Nov. 2002.
- [5] 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.
- [6] 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.
- [7] J.L.Shen, J.W.Hung and L.S.Lee, "Robust entropy-based endpoint detection for speech recognition in noisy environments", Proc.ICSLP-98, 1998.
- [8] 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.
- [9] C.Jia and B.Xu, "An improved entropy-based endpoint detection algorithm", Proc.Int.Sympo.Chinese Spoken Language Processing, pp.1399-1402, Aug.2002.
- [10] B.F.Wu and K.C.Wang, "Noise spectrum estimation with entropy-based VAD in non-stationary environments", IEICE Trans.Fundamentals, vol.E89-A, No2, Feb, 2006.
- [11] 佐々木潤子, 田中雅史 "マスキング効果を用いた低歪み雑音低減方式の検討", 信学技法, EA98-106, Dec. 1998.
- [12] 鈴木大和, 中山謙二, 平野晃宏 "スペクトルサブ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.
- [13] 川村新, 飯國洋二 "雑音追従性能を改善した MAP 推定による音声強調", 信学技法, SIP2006-19, May. 2006.
- [14] 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.
- [15] 鈴木大和, 中山謙二, 平野晃宏 "スペクトルサブプレッション法における雑音スペクトル推定の改善", 電子情報通信学会 2006 年ソサイエティ大会, 金沢, A-4-21, pp.88, Sep.2006.