

On Improvement of Voice Quality in Noise Cancelers Based on Spectrum Suppression Approach

メタデータ	言語: jpn 出版者: 公開日: 2017-10-03 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	http://hdl.handle.net/2297/18407

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

On Improvement of Voice Quality in Noise Cancelers Based on Spectrum Suppression Approach

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
[‡]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.

あらまし

本論文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の音声品質の改善について検討する。スペクトルサプレッション法を用いて、単一マイク方式のノイズキャンセラを実現する際には、雑音抑圧における音質の劣化と雑音抑圧によって新たに発生するミュージカルノイズを如何に抑えるかがポイントとなる。ミュージカルノイズを抑えるために、その周波数成分が主として高域にあることを利用して、高域抑圧を行なった。次に、雑音を過剰に抑圧するために音声に歪みが発生し音質が劣化する。この劣化を防ぐために、原音(観測信号=音声+雑音)をある割合だけ付加する方法を検討した。従来方法に比べて、ミュージカルノイズの抑制と音質の改善が得られた。

ABSTRACT

In this paper, some methods, which improve voice quality and suppress musical noise in noise cancellers based on spectrum suppression approach, have been discussed. First, in order to suppress musical noise, the spectrum is weighted so as to suppress the high frequency components, which are dominant in the musical noise. Second, in order to improve voice quality, the original noisy voice is added to the voice after noise cancellation. These methods have been simulated and compared to the conventional methods. The simulation results show the proposed methods can improve the voice quality and suppress the musical noise. How-

ever, the method adding the original noisy voice requires more proper threshold control.

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きている。そのため、街頭や車内など背景雑音の大き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。このような雑音を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。

現在、様々なノイズキャンセラの方式が提案されているが、ETSIが定める3GPP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、ノイズキャンセラとしては、スペクトルサプレッション法を採用する。スペクトルサプレッション法を用いて、単一マイク方式のノイズキャンセラを実現する際には、雑音抑圧における音声の劣化と雑音抑圧によって新たに発生するミュージカルノイズが問題となり、これらを抑えるかがスペクトル

サブプレッション法のポイントとなる．本稿では，スペクトルサブプレッション法における音質の改善とミュージカルノイズの抑制について検討を行う．

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．

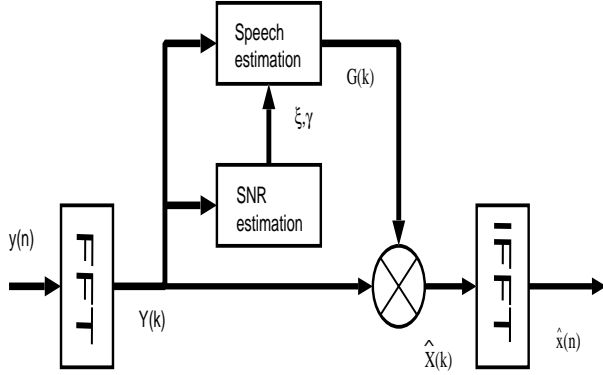


図 1: スペクトルサブプレッション法のブロック図

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平均 0 であるとする．時間領域でのクリアな音声を $x(n)$ ，雑音を $d(n)$ ，雑音混入音声を $y(n)$ とおくと，

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

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， n 番目のフレームにおける切り出された雑音混入音声 $\hat{y}_n(n)$ は，

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

と表せる．

この信号の周波数領域での表示を次のように表す．

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

$$Y(k) = R_k \exp(j\theta_k) \quad (4)$$

$$X(k) = A_k \exp(j\alpha_k) \quad (5)$$

雑音の分散を $\lambda_d(k)$ 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) はそれぞれ次の式で表せる．

$$\xi_n(k) = \frac{A_k^2}{\lambda_d(k)} \quad (6)$$

$$\gamma_n(k) = \frac{R_k^2}{\lambda_d(k)} \quad (7)$$

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 $\xi_n(k)$ は，次式で推定できる [1]．

$$\hat{\xi}_n(k) = \alpha \gamma_{n-1}(k) G_{n-1}^2(k) + (1 - \alpha) P[\gamma_n(k) - 1] \quad (8)$$

ただし，

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

また，雑音の分散は，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．重み付き雑音推定では，信号対雑音比 (SNR) 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．推定された雑音を基に，事後 SNR $\gamma_n(k)$ 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 $G_n(k)$ を求め，それを雑音混入音声に掛けることにより雑音を抑える．

2.2 MMSE STSA 法

MMSE STSA 法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 [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．スペクトルゲインは

$$G_n(k) = \frac{\left[(1 + \nu_n(k)) I_0\left(\frac{\nu_n(k)}{2}\right) + \nu_n(k) I_1\left(\frac{\nu_n(k)}{2}\right) \right] \cdot \frac{\Lambda_n(k)}{1 + \Lambda_n(k)} \frac{\sqrt{\pi \nu_n(k)}}{2 \gamma_n(k)} \exp\left(-\frac{\nu_n(k)}{2}\right)}{\quad} \quad (10)$$

で求められる．式中の各関数は

$$\nu_n(k) = \frac{\eta_n(k)}{1 + \eta_n(k)} \hat{\gamma}_n(k) \quad (11)$$

$$\Lambda_n(k) = \frac{1 - q_k}{q_k} \cdot \frac{\exp(\nu_n(k))}{1 + \nu_n(k)} \quad (12)$$

$$\eta_n(k) = \frac{\hat{\xi}_n(k)}{1 - q_k} \quad (13)$$

で求められる． I_0, I_1 はそれぞれ 0 次と 1 次の Bessel 関数， q_k は雑音混入音声中に音声成分が存在しない確率を表している．

2.3 Joint MAP 法

MMSE STSA 法は、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してスペクトルゲインを求めた。しかし、音声のスペクトル成分は、スーパーガウス分布であることが知られている。統計モデルとしてガウス分布とした場合は、その振幅はレーリー分布をとる [5]。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。

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

ただし、式中の A は音声の振幅を表していて、 τ と μ は変数であり、 $p(A)$ の形状を決定する。 $\tau = 0.01$ 、 $\mu = 1.5$ のとき最も音声をモデル化するのに適している [5]。

Joint MAP 法では、雑音混入音声 Y という条件での音声の振幅 A と偏角 α における条件付き確率 $p(A, \alpha | Y)$ を最大にする振幅 A と偏角 α を雑音混入音声 Y のクリア音声の推定振幅、偏角の値 \hat{A} 、 $\hat{\alpha}$ とする [5]。スペクトルゲインは、

$$G_{JMAP} = u + \sqrt{u^2 + \frac{\tau}{2\gamma}} \quad (15)$$

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

で求められる。

3 ミュージカルノイズの抑制と音質改善

3.1 ミュージカルノイズの抑制 –高域抑制–

スペクトルサプレッション法を用いてノイズキャンセラを行なった場合、雑音抑圧後にミュージカルノイズが発生する。そこで、ミュージカルノイズを抑えるためにミュージカルノイズの周波数帯域が高域であることから、図 2 のような重み $w(f)$ をスペクトルゲインに掛けることによる高域抑制を行なう。

$$w(f) = 0.5 \cos\left(2\pi \frac{f}{f_s}\right) + 0.5 \quad (17)$$

ただし、 f_s はサンプリング周波数である。

3.2 音質改善 –原音付加–

雑音抑圧の時に、抑圧のしすぎにより音声に歪みが発生し音質が劣化してしまう。そこで、雑音抑圧後音声に、事後 SNR (雑音混入音声対雑音比) $\gamma_n(k)$ に基づいて、原

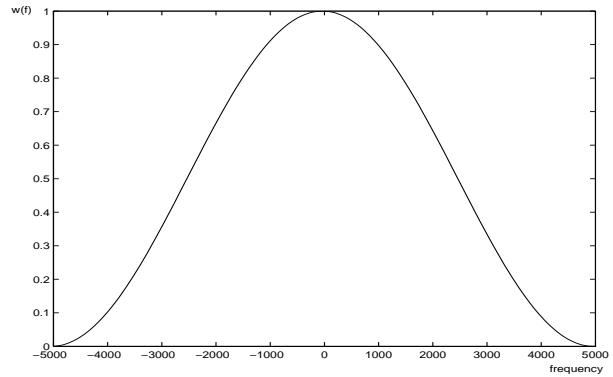


図 2: 重み $w(f)$ (高域抑制)

音 (観測信号 = 音声 + 雑音) をある割合だけ付加することにより雑音抑圧によって生じた音声の歪みを補う。

$$\hat{X}(k) = \begin{cases} \hat{X}(k) + \alpha_1 Y(k) & (2 < \gamma_n(k)) \\ \hat{X}(k) + \alpha_2 \gamma_n(k) Y(k) & (1 < \gamma_n(k) \leq 2) \\ \hat{X}(k) & (\text{otherwise}) \end{cases} \quad (18)$$

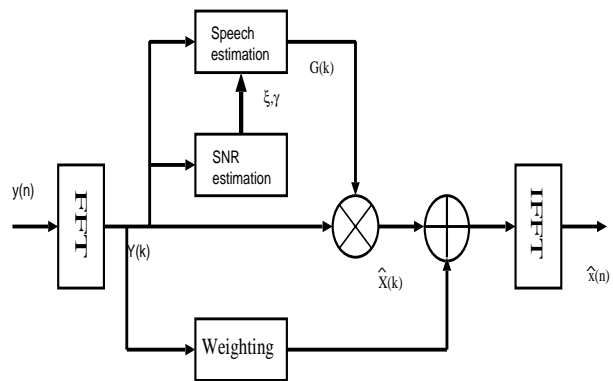


図 3: 原音付加

4 シミュレーション

文献 [2] が提案している重み付き雑音推定に基づく MMSE STSA 法によるノイズサプレッションを従来法とし、従来法に提案法を用いた場合についてシミュレーションを行なった。表 1 にシミュレーション条件を示す。

また、雑音抑圧後の評価として、客観評価として、信号対雑音比 (SNR) 評価を行ない、主観評価として、実際に音声を聞いてみて、ノイズサプレッション前とノイズサプレッション後の音質を比較した。

信号対雑音比 (SNR) 評価として、ノイズサプレッション前の信号対雑音比評価として、雑音の無い音声 $x(n)$

の二乗平均と雑音 $d(n)$ の二乗平均の比を $SNR1$ とする．ノイズサプレッション後の信号対雑音比評価として，雑音の無い音声 $x(n)$ の二乗平均と雑音の無い音声 $x(n)$ とノイズサプレッション後の音声 $\hat{x}(n)$ との誤差 $e(n)$ の二乗平均の比を $SNR2$ とする．それぞれ次式で表せる．

$$SNR1 = 10 \log_{10} \frac{|x(n)|^2}{|d(n)|^2} \quad (19)$$

$$e(n) = x(n) - \hat{x}(n) \quad (20)$$

$$SNR2 = 10 \log_{10} \frac{|x(n)|^2}{|e(n)|^2} \quad (21)$$

分析フレーム長	L=256 サンプルの Hamming 窓
オーバーラップ長	M=128 サンプル
評価方法	SNR, 主観評価
入力音声	サンプリング周波数 10kHz の男性の音声
雑音,SNR	白色雑音:約 6dB, バブル雑音:約 9dB

表 1: シミュレーション条件

4.1 MMSE STSA 法によるノイズサプレッション

4.1.1 従来法

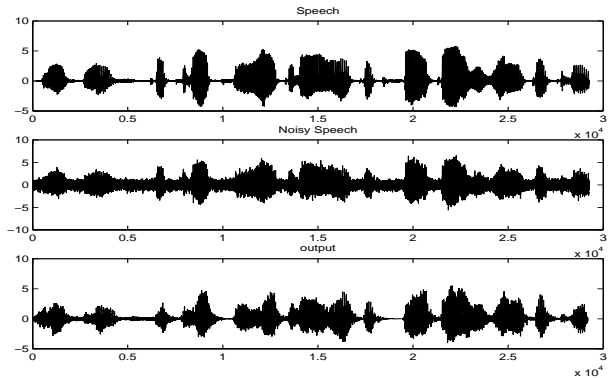


図 4: 従来法．白色雑音．上から順に雑音の無い音声/雑音混入音声/ノイズサプレッション後音声．SNR 改善量 ($SNR2 - SNR1$):-3.1138[dB]

図 4 に示す波形は，文献 [2] が提案している重み付き雑音推定に基づく MMSE STSA 法によるノイズサプレッションを行なった場合の，白色雑音の時の音声波形で，上から雑音の無い音声波形，雑音混入音声波形，ノイズサプレッション後の音声波形を表している．同様に，図 5 は，バブル雑音の時の音声波形を表している．どちらの場合も実際に雑音抑圧後音声を聞いてみたところ雑音はかなり消えているが音質が劣化してしまった．特に，バブル雑音のときの音質の劣化が大きかった．

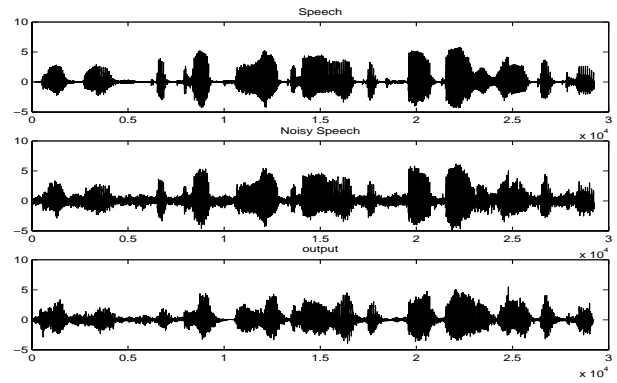


図 5: 従来法．バブル雑音．上から順に雑音の無い音声/雑音混入音声/ノイズサプレッション後音声．SNR 改善量 ($SNR2 - SNR1$):-0.4092[dB]

4.1.2 改善法 -原音付加，高域抑制-

従来法に音質改善 ($\alpha_1 = 0.3, \alpha_2 = 0.15$) を行なった場合の白色雑音の時の音声波形を図 6，バブル雑音の時の音声波形を図 7 に示す．どちらの場合も雑音抑圧後音声を聞いてみたところ音質が従来法よりも改善した．しかし，従来法よりも雑音が少し残る．また，バブル雑音のとき，雑音抑圧後の音声はこもったように聞こえる．

4.2 Joint MAP 法によるノイズサプレッション

4.2.1 重み付き雑音推定に基づく Joint MAP 法

文献 [2] が提案している重み付き雑音推定を用い，Joint MAP 法 [5] によるノイズサプレッションを行なった場合の，白色雑音の時の音声波形を図 8 に，バブル雑音の時の音声波形を図 9 に示す．どちらの場合も MMSE STSA 法よりもノイズを抑えることができ音質も MMSE STSA 法と比べるとそれほど劣化しない．しかし，音質はこもった感じに聞こえる．

4.3 Joint MAP 法によるノイズサプレッションの音質改善

Joint MAP 法において式 (18) のように閾値を設定して原音付加を行ない，雑音抑圧音声を聞いてみたところあまり音質の改善がなかった．図 8, 図 9 より，音声が雑音より弱いところでは雑音抑圧後に音声がほぼ完全に抑えられてしまう．そこで，式 (22) のように，閾値を設定せず原音を付加する割合を固定してシミュレーションを行なった．

$$\hat{X}(k) = \hat{X}(k) + 0.15Y(k) \quad (22)$$

原音を付加する割合を固定したところ、雑音は残るが白色雑音、バブル雑音両方の場合で音質が今回シミュレーションを行なった中で最も改善した。特に、バブル雑音の場合音質が大きく改善した。しかし、白色雑音の場合雑音がやや大きくなってしまった。

白色雑音とバブル雑音の双方に適用できる方式として、雑音スペクトルの形や時間的な変化により雑音の特性を推定し、原音付加方式のパラメータを制御する方式が望ましい。

5 まとめ

本稿では、スペクトルサブレーション法による単一マイクノイズキャンセラにおいて、ミュージカルノイズを抑制する方法と音質を改善する方法について検討を行った。まず、ミュージカルノイズを抑えるために、その周波数成分が高域にあることを利用して、高域抑制を行なう方法を検討した。次に、雑音抑圧後の音声に原音(観測信号=音声+雑音)をある割合で付加することにより雑音抑圧によって生じた音声の歪みを補う方法について検討した。

さらに、従来法 [2] と本稿で改良を加えた方法についてシミュレーションを行ない、従来法では雑音をかなり抑えることができるが音質が劣化すること、特に、音声とスペクトルや非定常性が似ているバブル雑音の場合劣化が大きいことを確認した。また本稿で改良を加えた方法を用いると、雑音は少し残るが、従来法より音質がかなり改善されることを確認した。しかし、白色雑音のときは雑音がやや大きくなってしまった。

今後の課題として、雑音の有色性や非定常性を評価し、音源付加を制御する閾値を可変にする必要がある。また、本稿では、白色雑音とバブル雑音のみで検討を行なったので、他の種々の雑音を用いて有効性について検証する必要がある。

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", in Proc. of EUSIPCO-04(Vienna, Austria), pp.1447-60, Sep. 2004. 金沢大学 工学部 電気・情報工学科 卒業論文, Feb. 2002. 動作解析, 金沢大学 工学部 電気・情報工学科 卒業論文, Feb. 2004.

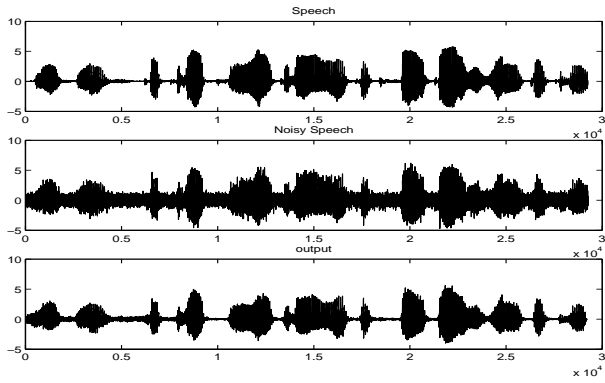


図 6: 音質改善を行なった MMSE STSA 法 . 白色雑音 . 上から順に雑音の無い音声/雑音混入音声/ノイズサブプレッション後音声 . SNR 改善量:4.5806[dB]

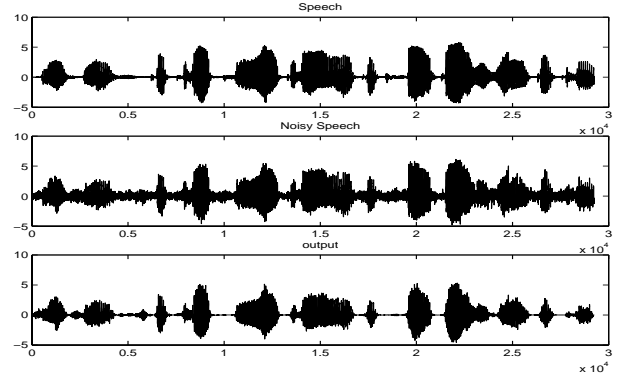


図 9: 重み付き雑音推定に基づく Joint MAP 法 . バブル雑音 . 上から順に雑音の無い音声/雑音混入音声/ノイズサブプレッション後音声 . SNR 改善量:-1.4019[dB]

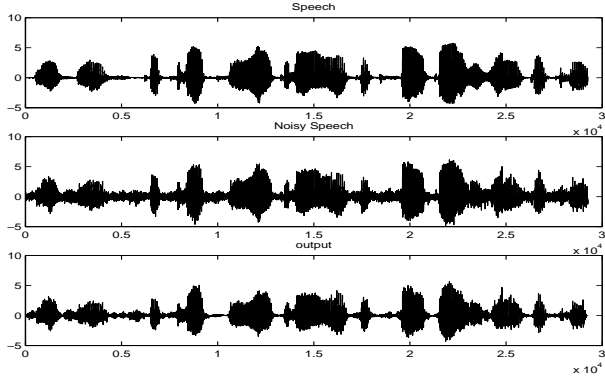


図 7: 音質改善を行なった MMSE STSA 法 . バブル雑音 . 上から順に雑音の無い音声/雑音混入音声/ノイズサブプレッション後音声 . SNR 改善量:-0.4092[dB]

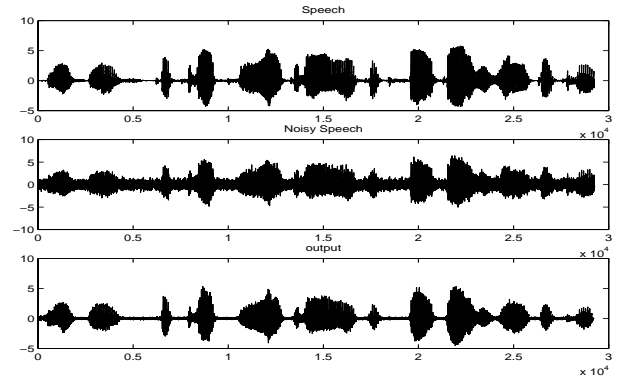


図 10: 原音付加(固定)を行なった Joint MAP 法 . 白色雑音 . 上から順に雑音の無い音声/雑音混入音声/ノイズサブプレッション後音声 . SNR 改善量:3.5863[dB]

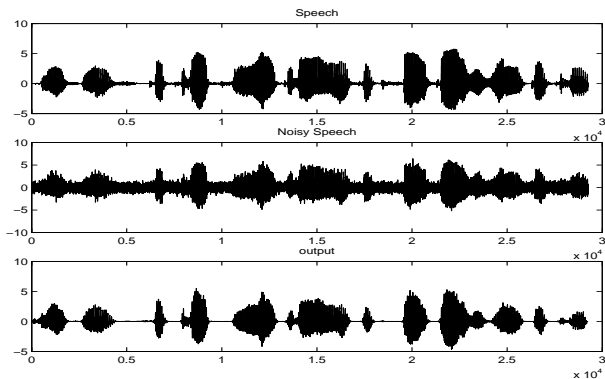


図 8: 重み付き雑音推定に基づく Joint MAP 法 . 白色雑音 . 上から順に雑音の無い音声/雑音混入音声/ノイズサブプレッション後音声 . SNR 改善量:2.6823[dB]

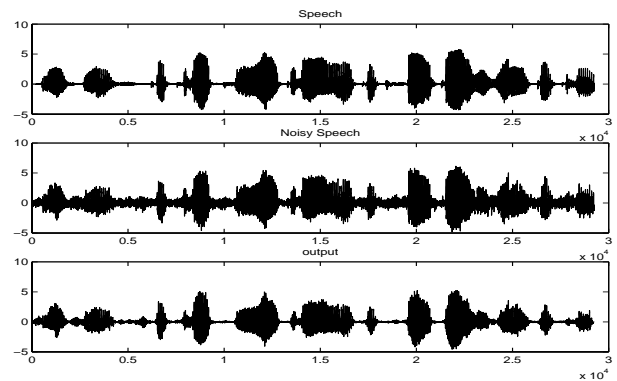


図 11: 原音付加(固定)を行なった Joint MAP 法 . バブル雑音 . 上から順に雑音の無い音声/雑音混入音声/ノイズサブプレッション後音声 . SNR 改善量:-0.4612[dB]