

Title	変調伝達関数に基づいた雑音抑圧に関する研究
Author(s)	山崎, 悠
Citation	
Issue Date	2009-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/8099
Rights	
Description	Supervisor: 鷓木 祐史, 情報科学研究科, 修士

修 士 論 文

変調伝達関数に基づいた雑音抑圧に関する研究

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

山崎 悠

2009年3月

修 士 論 文

変調伝達関数に基づいた雑音抑圧に関する研究

指導教官 鵜木 祐史 准教授

審査委員主査 鵜木 祐史 准教授
審査委員 赤木 正人 教授
審査委員 徳田 功 准教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

0710073 山崎 悠

提出年月: 2009 年 2 月

概要

我々は普段、色々な音が混在する実環境で生活している。この実環境下で、人々は音を聴くことによって、重要な情報を得ている。例えば、音楽を聴くことで心を豊かにしたり、音声を発声、聴いたりすることによって、コミュニケーションを取っている。しかし、実環境で観測される音楽や音声信号は、雑音や残響が混在した状態で観測される。雑音や残響によって、音声信号は歪み、音質や音声明瞭度が低下する。そのため、雑音や残響の影響を取り除くことは、音声認識システムや補聴システムの前処理といった音声信号処理で重要な研究課題である。

これまでの研究は、雑音環境下または残響環境下で取り組まれ、数多くの手法が提案されている。代表的な手法として、雑音環境下では、雑音成分を周波数領域で差し引く Boll のスペクトルサブトラクション法や Paliwal と Basu のカルマンフィルタリング、残響環境下では、室内伝達特性の最小位相成分を取り除く Neely と Allen の手法やマイクロホンアレーを用いて、逆フィルタリングをする三好と金田の MINT 法が知られている。しかしながら、雑音や残響が混在する実環境では、これらの手法が、有効に機能するとは考えにくい。最近になり、雑音・残響環境下での雑音・残響抑圧法も提案されている。木下らは、雑音に対してはスペクトル減算を残響に対しては線形予測を用いて、縦列的に抑圧する手法を提案した。しかし、雑音と残響を同時に抑圧するためには、雑音・残響環境下で共通に扱える特徴あるいは特性が必要である。

MTF は、信号の伝送路の特徴を時間強度包絡線の変調度で関係づけるものである。Houtgast と Steeneken は、MTF を利用して音声明瞭度を予測する体系を提案した。この音声明瞭度は、人間がコミュニケーションを取る上で非常に重要な要素である。この理論体系は、雑音・残響による音声明瞭度の低下を考慮している。そのため、MTF を用いることで、音声明瞭度を考慮した雑音・残響抑圧法の提案が期待できる。

鷓木らは、MTF に基づく残響抑圧法を提案した。彼らの手法は、残響環境下に限定して回復処理を行なっている。この手法は残響によって低下した音声明瞭度を約 30% ほど改善することができる。この手法に MTF に基づく雑音抑圧法を組み込むことにより、雑音・残響環境下での MTF に基づいた音声回復処理法の提案が望める。

本研究では、雑音・残響環境での人間の円滑なコミュニケーションの達成を目指した雑音・残響抑圧法の提案を最終目標とし、音声明瞭度を考慮した MTF に基づく雑音抑圧法を提案する。雑音の影響を受けた入力パワーエンベロープは、変調度だけでなく振幅も影響を受けているため、変調度に関係する MTF を回復するだけでなく振幅も回復することで、観測パワーエンベロープから入力パワーエンベロープを得る。振幅に関しては、振幅補正値を掛けることで回復する。変調度に関しては、MTF の逆数 (IMTF) を掛けることで回復する。これにより雑音の抑圧を行なう。

IMTF を求めるには MTF を算出しなければならない。雑音環境下での MTF を算出するには、入力パワーエンベロープの平均値が必要となるため、この平均値の推定を行なう必

要がある．まず無音声区間から，雑音パワーエンベロープの平均値を得る．そして，音声区間の平均値から，求めた雑音パワーエンベロープの平均値を引くことで，入力パワーエンベロープの平均値を算出し，MTFを推定した．

提案法を評価するために，評価シミュレーションを行った．評価に用いる音声は，ATRデータベースにある男性5名，女性5名，計10名の話者が発話した3単語とした．SNRが20，10，5，0，-5 dBになるように白色，ピンク，バブル雑音を付加した．1つのSNRに対して雑音をそれぞれ100個用意した．評価項目として，相関値，パワー比の改善度，対数スペクトル距離(LSD)と音声明瞭度と関係の取れた重み付きLSDを用いた．その結果，パワー比の改善度ではSNRが低くなるごとに増加した．またLSDでは最大約31 dBほど，重み付きLSDでは最大約8 dBの改善が見られた．相関値は大きな違いは認められなかった．以上のことからMTFに基づく雑音抑圧法が時間強度包絡線と信号回復の点で，雑音音声の回復に寄与していることを示した．

目次

第1章	序論	1
1.1	背景	1
1.2	目的	2
1.3	本論文の構成	2
第2章	従来 of 雑音抑圧法	4
第3章	変調伝達関数 (MTF)	7
3.1	MTF の原理 (概念)	7
3.2	雑音環境での MTF	7
3.3	残響環境での MTF	8
3.4	雑音・残響環境での MTF	10
第4章	MTF に基づいたパワーエンベロープ逆フィルタ法	11
4.1	信号の生成過程	11
4.2	信号とパワーエンベロープの関係	11
4.3	パワーエンベロープの抽出法	14
4.4	MTF に基づくパワーエンベロープ逆フィルタ法	14
第5章	提案法	15
5.1	雑音・残響抑圧の処理体系	15
5.2	雑音抑圧法を提案する際の問題点	15
5.3	提案法のコンセプト	15
5.4	MTF に基づいた雑音抑圧法	16
5.4.1	MTF に基づいた雑音抑圧法の原理	16
5.4.2	雑音環境での MTF の推定	17
5.5	音声信号への適応と帯域分割処理	19
5.6	パワーエンベロープを差し引く手法	22
第6章	提案法の評価	23
6.1	条件	23
6.2	結果と考察	24

6.2.1	相関, SNR の改善度による評価	24
6.2.2	LSD による評価	25
6.2.3	重み付け LSD による評価	25
第 7 章	結論	44
7.1	本研究で明らかにしたこと	44
7.2	本研究における今後の課題と展望	44

目次

2.1	入力音声 $x(t)$, 観測音声 $y(t)$ と従来法を施した音声 (SNR 0 dB, 音声は/aikawarazu/)	6
3.1	雑音環境下での MTF	8
3.2	残響環境下での MTF	9
3.3	雑音・残響環境下での MTF	10
4.1	伝達関数の構成 (a) は信号の場合, (b) はパワーエンベロープの場合	12
4.2	信号とパワーエンベロープの関係	13
5.1	MTF に基づいた雑音抑圧法の構成	16
5.2	パワーエンベロープでの MTF に基づいた雑音抑圧法の流れ, (a) から (e) の実線は各式でのパワーエンベロープの状態, (f) は観測パワーエンベロープに $1/m(f_m)$ 倍した場合, 破線は, 入力パワーエンベロープ	18
5.3	入力音声 $x(t)$, 観測音声 $y(t)$ と提案法を施した音声 (SNR 0 dB, 音声は/aikawarazu/)	20
5.4	提案法のブロック図	21
6.1	提案法:白色雑音での相関値とパワー比の改善度	26
6.2	提案法:ピンク雑音での相関値とパワー比の改善度	27
6.3	提案法:バブル雑音での相関値とパワー比の改善度	28
6.4	ウィナーフィルタ:白色雑音での相関値とパワー比の改善度	29
6.5	ウィナーフィルタ:ピンク雑音での相関値とパワー比の改善度	30
6.6	ウィナーフィルタ:バブル雑音での相関値とパワー比の改善度	31
6.7	SS 法:白色雑音での相関値とパワー比の改善度	32
6.8	SS 法:ピンク雑音での相関値とパワー比の改善度	33
6.9	SS 法:バブル雑音での相関値とパワー比の改善度	34
6.10	MMSE-STSA:白色雑音での相関値とパワー比の改善度	35
6.11	MMSE-STSA:ピンク雑音での相関値とパワー比の改善度	36
6.12	MMSE-STSA:バブル雑音での相関値とパワー比の改善度	37
6.13	白色雑音での LSD の改善度	38
6.14	ピンク雑音での LSD の改善度	39
6.15	バブル雑音での LSD の改善度	40
6.16	白色雑音での重み付き LSD の改善度	41
6.17	ピンク雑音での重み付き LSD の改善度	42

6.18 バブル雑音での重み付き LSD の改善度	43
-------------------------------------	----

第1章 序論

1.1 背景

我々は普段、色々な音が混在する実環境で生活している。この実環境下で、人々は音を聴くことによって、重要な情報を得ている。例えば、音楽を聴くことで心を豊かにしたり、音声を発声、聴いたりすることによって、コミュニケーションを取っている。このように、音を聴くということは我々の生活において重要である。しかし、実環境で観測される音楽や音声信号は、雑音や残響が混在した状態で観測される。雑音や残響によって、音声信号は歪み、音質や音声明瞭度が低下する。これにより、円滑なコミュニケーションが阻害されたり、音声認識システムや補聴システムの処理精度が低下する。これらを実環境で、円滑または精度良く行なうためには、雑音や残響の影響を取り除く必要がある。雑音や残響抑圧の研究は、これまでに数多く行なわれてきており、現在でも、音声信号処理で最も重要な課題の一つとして、研究が行なわれ続けている。

これまでの研究は、雑音環境下または残響環境下のいずれかで取り組まれ、数多くの手法が提案されている。代表的な手法として、残響環境下では、室内伝達特性の最小位相成分を取り除く Neely と Allen の手法 [1] やマイクロホンアレーを用いて、逆フィルタリングをする Miyoshi と Kaneda の MINT 法 [2] や Wang と Itakura の回復信号と音声信号の誤差を最小にする帯域分割逆フィルタ [3] が知られている。雑音環境下では、雑音成分を周波数領域で差し引く Boll の SS 法 [4] やウィナーフィルタ [5] を用いた雑音抑圧法 [6][7]、カルマンフィルタリング [8] による Paliwal と Basu の抑圧法 [9]、Ephraim と Malah による推定短時間振幅スペクトルの平均 2 乗誤差を最小にする MMSE-STSA 法 [10]、McAulay と Malpass の音声振幅スペクトルの最尤推定量を使用した ML 法 [11] がある。これらの手法は音声認識システムの前処理 [12][13] にも用いられており、近年でもその改良法 [14][15] が提案され続けている。しかし、雑音と残響が混在する実環境を考慮した場合、これらの雑音や残響抑圧法が、有効に機能するとは考えにくい。

最近になり、雑音・残響環境下での抑圧法も検討、提案されている。吉田らは雑音・残響環境で荒井らの定常部抑圧法 [16] の評価をしている [17]。彼らによると、定常部抑圧法により、雑音・残響環境で音声明瞭度の保持ができると述べている。しかし、定常部抑圧法は、呈示する前の音声に処理を施す手法である。そのため、実環境で観測した音声を処理する補聴システムや音声認識システムの前処理とするものではない。雑音・残響環境で音声を処理する手法として、Kinoshita らは雑音に対してはスペクトル減算を残響に対しては線形予測を用いて、縦列的に抑圧する手法を提案した [18]。また、吉岡らは雑音抑

圧と残響抑圧を協調して行なうために、観測音声のモデルを周波数領域で設定し、最尤推定法で求めたパラメータを用いて音声の回復を行なう手法 [19] とその改良法 [20] を提案している。これらの手法は雑音と残響を同時に抑圧する手法ではない。同時に雑音と残響を抑圧するためには、雑音・残響環境下で共通に扱える特徴あるいは特性が必要である。

変調伝達関数 (MTF : Modulation Transfer Function) は、信号の伝送路の特徴をパワーエンベロープの変調度で関係づけるものである。Houtgast と Steeneken は、MTF を利用して音声明瞭度 [21] を予測する体系 [22]-[24] を提案した。この理論体系は、MTF から音声明瞭度 (STI : Speech Transmission Index)[25][26] に変換することで、音声明瞭度と直接関係がとられており、雑音環境のみ残響環境のみだけではなく、雑音・残響環境での音声明瞭度の低下を考慮している。そのため、雑音・残響環境を考慮している MTF を用いることで、雑音・残響の同時抑圧法を提案できる可能性がある。

MTF に基づいて残響時間のブラインド推定 [27] や残響環境での音声認識 [28] が行なわれている。広林らは、パワーエンベロープの逆フィルタリングを行なうことで残響抑圧を行なった [29]。Unoki らは、広林らの手法を再検討し、新たに MTF に基づく残響抑圧法 [31]-[33] を提案した。彼らの手法は、残響環境下に限定して回復処理を行なっている。Unoki らの手法は残響によって低下した音声明瞭度を約 30% ほど改善することができる [33]。ここで Unoki らの手法に、MTF に基づく雑音抑圧法を組み込むことができれば、雑音・残響環境下での MTF に基づいた音声回復処理法を提案することができる。

1.2 目的

著者の目標は、雑音と残響を同時に抑圧可能な音声回復法を提案することである。そのため、本研究では、雑音により低下した MTF を復元することで、MTF に基づく雑音抑圧法を提案する。雑音の影響を受けた入力パワーエンベロープは、変調度だけでなく振幅も影響を受けているため、変調度に関係する MTF を回復するだけでなく振幅も回復することで、観測パワーエンベロープから入力パワーエンベロープを得る。この手法は、雑音抑圧法ではあるが、従来の雑音抑圧法とは違い、MTF に基づいていることにより、雑音だけでなく残響についても考慮できる可能性がある。

1.3 本論文の構成

本論文は、全 7 章により構成されている。以下に各章の概要を述べる。第 1 章では、本研究の背景である、残響や雑音に関する抑圧法と目的について述べる。第 2 章では、従来の雑音抑圧法について簡潔に述べる。第 3 章では、本研究で重要な概念である MTF について説明する。提案法で用いる雑音環境での MTF について述べ、MTF が雑音・残響環境を考慮していることを示す。第 4 章では、提案法で用いる、MTF に基づいた伝達系、信号とパワーエンベロープの関係とパワーエンベロープの抽出法について述べ、MTF に基

づいた雑音抑圧法を提案する際に指針となったパワーエンベロープの逆フィルタ法について述べる．第5章では，雑音・残響抑圧法の処理体系についてのコンセプトを述べる．また，雑音抑圧法を提案する際の問題点を明らかにしたうえで解決し，MTFに基づいた雑音抑圧法を提案する．第6章では，提案法の評価を行なう．第7章では，本研究で明らかにしたこと，本研究の今後の課題と展望について述べる．

第2章 従来の雑音抑圧法

従来の雑音抑圧法について，簡単に述べる．雑音環境での手法なので，信号の伝達系は， $y(t) = x(t) + n(t)$ の関係を用いる．そのスペクトルは $Y(\omega) = X(\omega) + N(\omega)$ となる．また，推定入力スペクトルは

$$\hat{X}(\omega) = G(\omega)Y(\omega) \quad (2.1)$$

$$G(\omega) = 1 - \frac{N(\omega)}{Y(\omega)} \quad (2.2)$$

となる．ここで $G(\omega)$ はゲイン関数である．ただし，式 (2.2) は理想的なゲイン関数である．これから説明する手法は，着目するスペクトルとゲイン関数の求め方が違えど，式 (2.1) により，入力情報を得る．

ウィナーフィルタリング

ウィナーフィルタリングは Wiener が提案したウィナーフィルタを用いて，雑音抑圧を行なう手法で，音声信号と推定した音声信号のパワースペクトルの平均2乗誤差を最小とする手法である．パワースペクトルの関係とゲイン関数は，

$$|Y(\omega)|^2 = |X(\omega)|^2 + |N(\omega)|^2 \quad (2.3)$$

$$|\hat{X}(\omega)|^2 = G_{WF}(\omega)|Y(\omega)|^2 \quad (2.4)$$

$$G_{WF}(\omega) = \frac{|\hat{X}(\omega)|^2}{|\hat{X}(\omega)|^2 + |N(\omega)|^2} \quad (2.5)$$

となる．ここで，精度良くフィルタリングが行われるためには，入力信号と雑音信号が無相関であることが重要である．

Spectral Subtraction(SS) 法

SS 法は観測信号の振幅スペクトルから雑音の振幅スペクトルの推定平均値を差し引くことで入力振幅スペクトルを得る手法である．振幅スペクトルの関係とゲイン関数は

$$|Y(\omega)| = |X(\omega)| + |N(\omega)| \quad (2.6)$$

$$|\hat{X}(\omega)| = 1 - \frac{G_{SS}(\omega)}{|Y(\omega)|} \quad (2.7)$$

$$G_{SS}(\omega) = 1 - \frac{|\hat{N}(\omega)|}{|Y(\omega)|} \quad (2.8)$$

となる．また，SS法は $|\hat{N}(\omega)|$ を $|Y(\omega)|$ から差し引くことでも $|\hat{X}(\omega)|$ を得ることができる．

MMSE-STSA 法

MMSE-STSA法はウィナーフィルタリングの様に，平均2乗誤差を最小にする手法である．しかし，パワースペクトルではなく，音声信号と推定した音声信号の振幅スペクトルであり，位相についても考慮されている．振幅スペクトルの関係とゲイン関数は，

$$|Y(\omega)| = |X(\omega)| + |N(\omega)| \quad (2.9)$$

$$\hat{X}(\omega) = G_{MMSE}(\omega)|Y(\omega)| \cdot \exp\{j\angle Y(\omega)\} \quad (2.10)$$

$$G_{MMSE}(\omega) = \frac{(\pi v)^{1/2}}{2\gamma} \exp\left(\frac{-v}{2}\right) \left[(1+v)I_0\left(\frac{v}{2}\right) + vI_1\left(\frac{v}{2}\right) \right] \quad (2.11)$$

となる．ここで， $I_0(\cdot)$ と $I_1(\cdot)$ は変形ベッセル関数を， $\angle Y(\omega)$ は位相スペクトルである．また，

$$v = \frac{\xi}{1 + \xi} \gamma, \quad \xi = \frac{\sigma_x^2}{\sigma_n^2}, \quad \gamma = \frac{|Y(\omega)|^2}{\sigma_n^2} \quad (2.12)$$

である．ここで， σ_x^2, σ_n^2 は入力スペクトルと雑音スペクトルの分散， ξ は事前SNR， γ は事後SNRである．以上が比較に用いる従来法の簡単な説明である．詳細について，SS法は文献[4]を，ウィナーフィルタリングは文献[7]，MMSE-STSA法は文献[10][14]に記述されている．図(2.1)は入力音声(/aikawarazu/)に，SNR 0 dBとなるように雑音を加算した観測音声と前述の雑音抑圧法を観測音声に施した結果の信号波形である．図から，雑音が抑圧されているのが見て取れる．

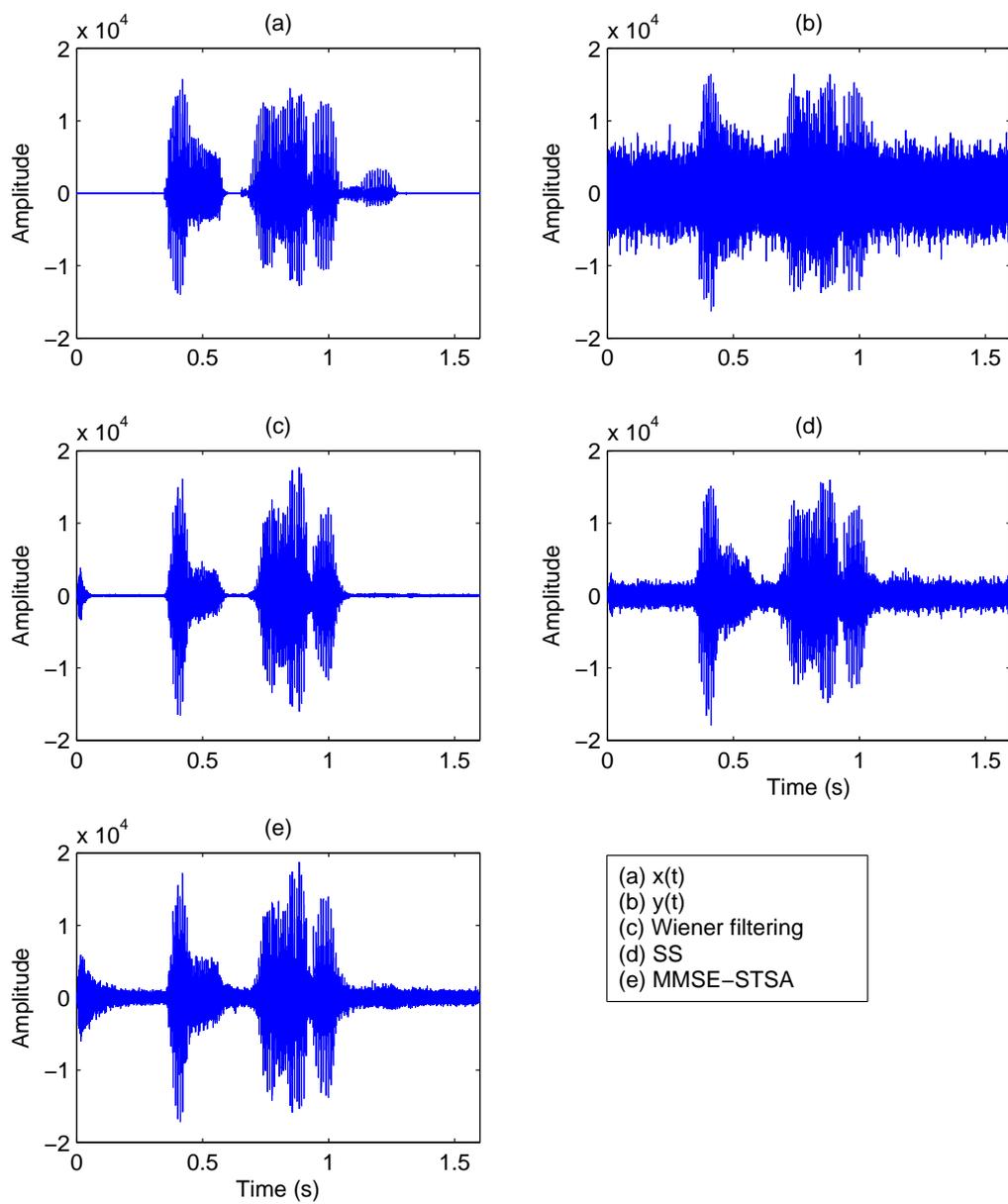


図 2.1: 入力音声 $x(t)$, 観測音声 $y(t)$ と従来法を施した音声 (SNR 0 dB, 音声は/aikawarazu/)

第3章 変調伝達関数(MTF)

MTF の概念は、入出力信号のパワーエンベロープの変調度と伝達系の特性の関係を説明するために Houtgast と Steeneken によって提案された。この概念は、音声明瞭度への伝達系の効果を評価するための概念として室内音響学において紹介された。

3.1 MTF の原理 (概念)

Houtgast と Steeneken は、入出力のパワーエンベロープを

$$\text{Input} = \overline{I}_i^2(1 + \cos(2\pi f_m t)), \quad (3.1)$$

$$\text{Output} = \overline{I}_o^2\{1 + m(f_m) \cos(2\pi f_m(t - \tau))\}, \quad (3.2)$$

と定義した [22]-[24]。ここで、 \overline{I}_i^2 と \overline{I}_o^2 は、入出力の強度、 f_m は変調周波数、 τ は位相を表す。 $m(f_m)$ は変調周波数 f_m の変調度である。この $m(f_m)$ が雑音や残響の影響を受けることで変化し、この変化が変調度に相当することから MTF と呼ばれる。MTF は STI に変換されることで、音声明瞭度と直接関係が持たれる。

3.2 雑音環境での MTF

雑音環境での MTF について説明する。入力パワーエンベロープ $e_x^2(t)$ を

$$e_x^2(t) = \overline{e}_x^2(1 + \cos(2\pi f_m t)) \quad (3.3)$$

とすると、雑音を加算された場合の観測パワーエンベロープである $e_y^2(t)$ は、

$$e_y^2(t) = \overline{e}_x^2\{1 + \cos(2\pi f_m t)\} + e_n^2(t) \quad (3.4)$$

$$= (\overline{e}_x^2 + \overline{e}_n^2)\{1 + m(f_m) \cos(2\pi f_m t)\} \quad (3.5)$$

となる。ここで $e_n^2(t)$ は雑音信号のパワーエンベロープである。但し、 $e_n^2(t)$ は時間に一定と仮定されているため、 $e_n^2(t) = \frac{1}{T} \int_0^T e_n^2(t) dt$ である。ここで T は信号長である。また、雑音環境での MTF、 $m(f_m)$ は下記式で与えられる。

$$m(f_m) = \frac{\overline{e}_x^2}{\overline{e}_x^2 + \overline{e}_n^2} = \frac{1}{1 + 10^{-(\text{SNR})/10}}, \quad (3.6)$$

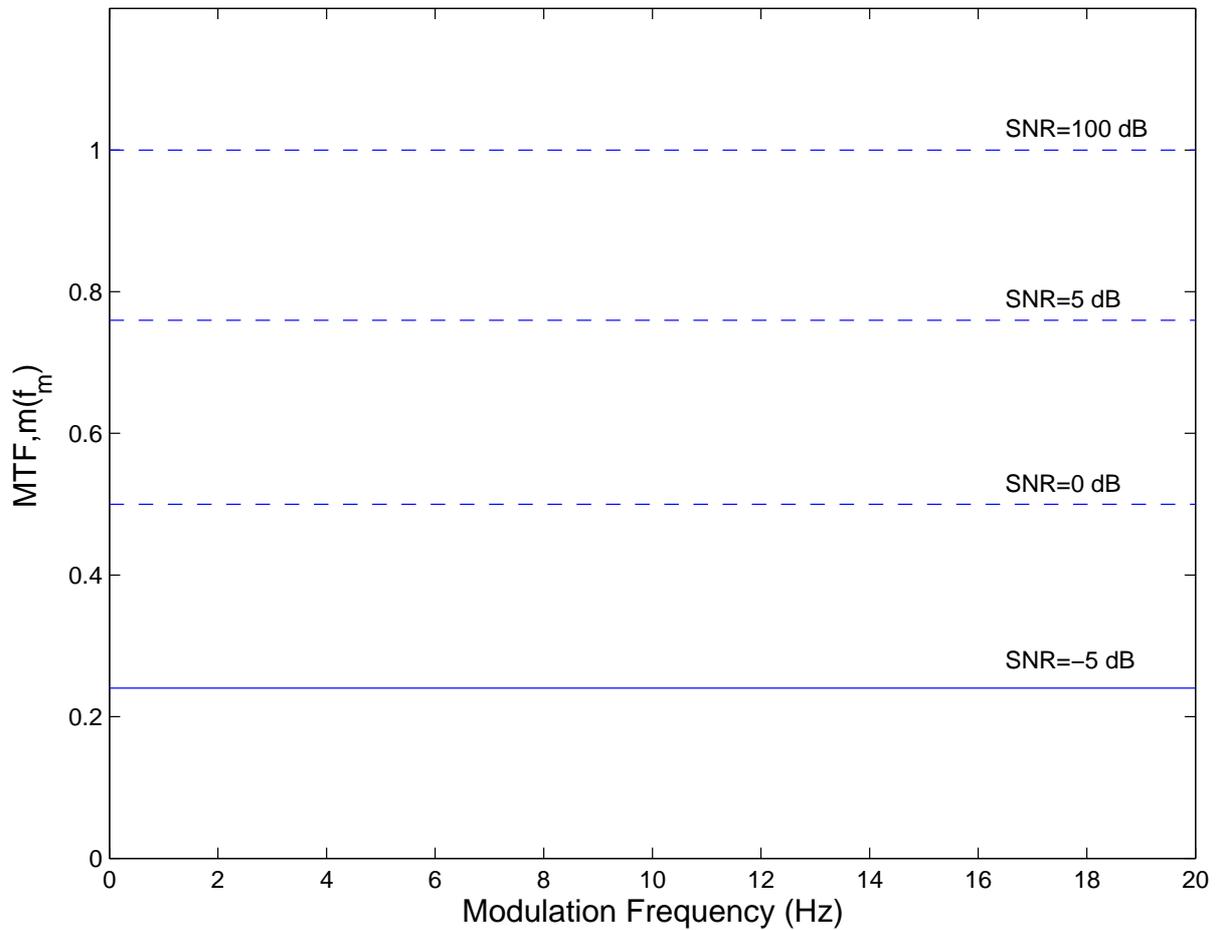


図 3.1: 雑音環境下での MTF

但し, $\text{SNR} = 10 \log_{10}(\overline{e_x^2(t)}/\overline{e_n^2(t)})$ dB である. 雑音環境での MTF は, f_m に依存しない. 例えば, SNR = -5 dB の場合, $m(f_m) \doteq 0.24$ である. (図 3.1)

3.3 残響環境での MTF

残響環境での MTF は, 下記式で与えられている.

$$m(f_m) = \frac{|\int_0^\infty h^2(t) \exp(-j2\pi f_m t) dt|}{\int_0^\infty h^2(t) dt} \quad (3.7)$$

ここで $\mathbf{h}(t)$ は室内インパルス応答であり, Schroeder の確率論的近似インパルス応答 [34] を用いれば, $\mathbf{h}(t)$ は

$$\mathbf{h}(t) = \exp\left(-\frac{6.9t}{T_R}\right) \cdot n_h(t) \quad (3.8)$$

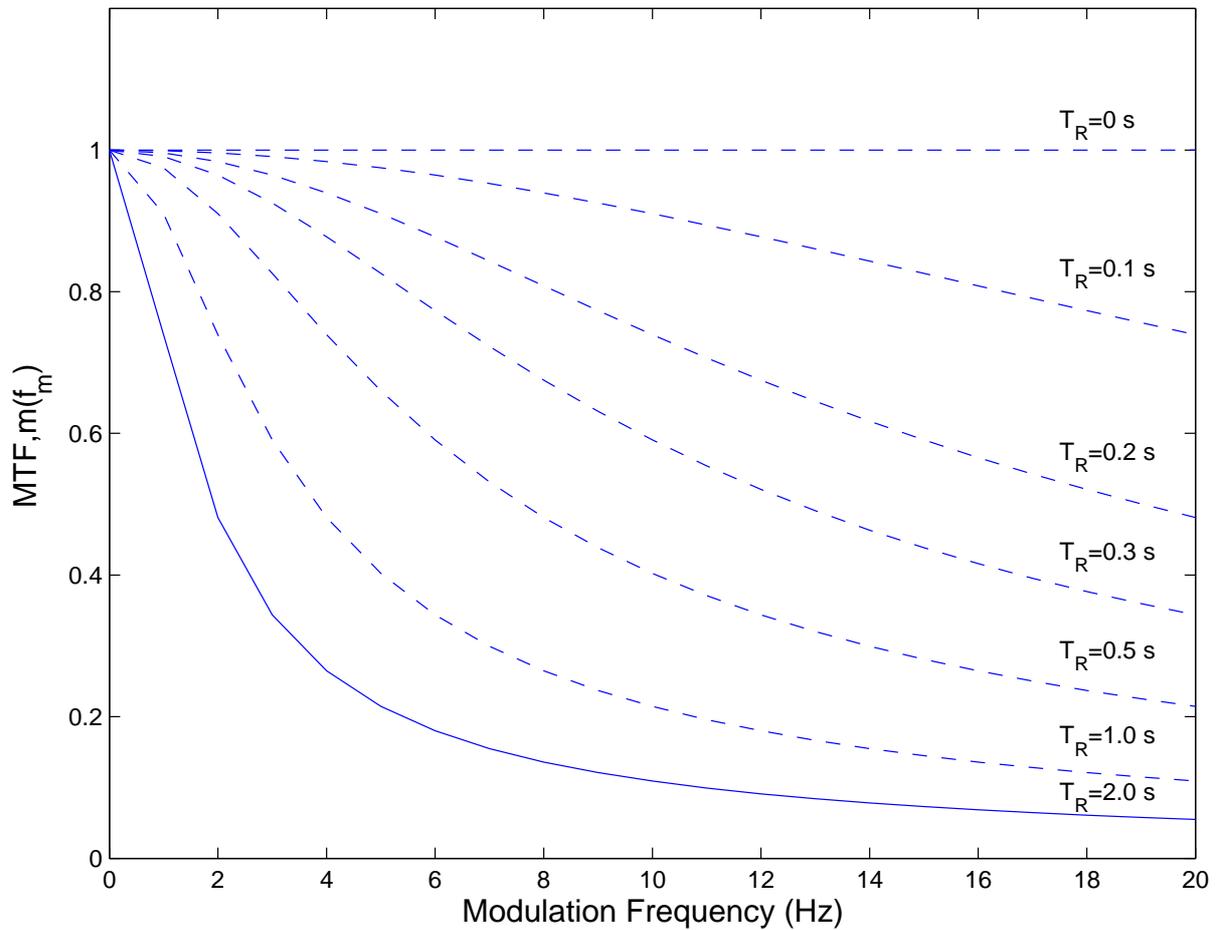


図 3.2: 残響環境下での MTF

と表すことができる．ここで T_R はパワーが 60 dB 減衰するまでの時間，残響時間である． $n_h(t)$ はキャリアでランダム変数である．式 (3.8) を式 (3.7) に代入すると，残響環境での MTF を下記式で表すことができる．

$$m(f_m) = \left[1 + \left(2\pi f_m \frac{T_R}{13.8} \right)^2 \right]^{-1/2} \quad (3.9)$$

この式 (3.9) と図 (3.2) より，残響環境での MTF は，変調周波数 f_m と残響時間 T_R に依存しており，その特性はローパス特性を示している．

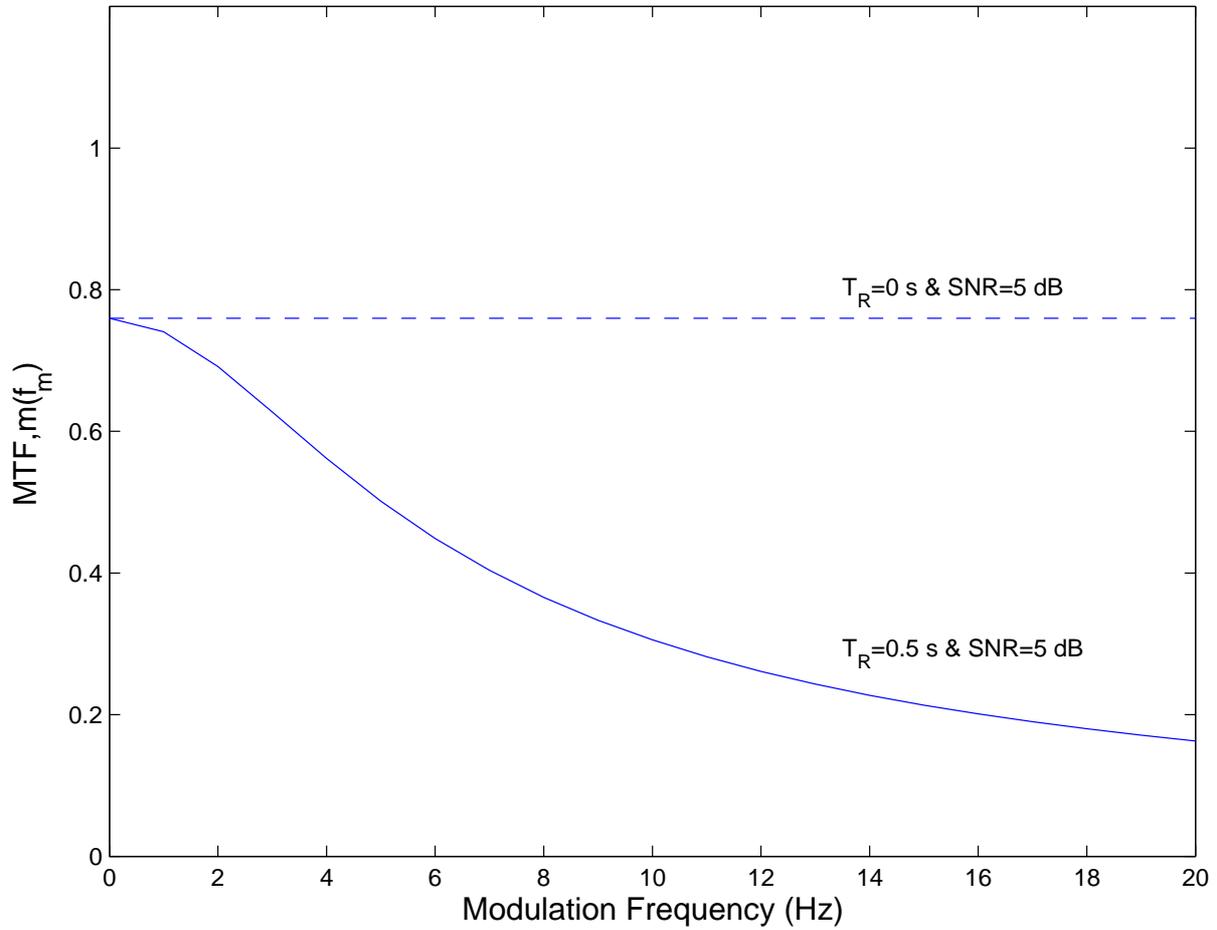


図 3.3: 雑音・残響環境下での MTF

3.4 雑音・残響環境での MTF

雑音・残響環境での MTF は，式 (3.1)，(3.2) の関係と式 (3.6)，(3.7) より下記のように表せれる．

$$m(f_m) = \left[1 + \left(2\pi f_m \frac{T_R}{13.8} \right)^2 \right]^{-1/2} \times \left(1 + 10^{-(\text{SNR})/10} \right)^{-1}. \quad (3.10)$$

このように，MTF に着目すれば，雑音・残響の影響を同時に考慮することができる．図 (3.3) は雑音・残響環境での MTF である．残響環境時の MTF と同じ様に指数減衰していくが，雑音の影響分だけ低下した状態から減衰している．本研究では，雑音環境を想定しているため，雑音環境での MTF に特化する．

第4章 MTFに基づいたパワーエンベロープ逆フィルタ法

4.1 信号の生成過程

本研究では，MTFに基づいて，観測信号を $y(t)$ ，入力信号を $x(t)$ ，インパルス応答を $h(t)$ ，雑音を $n(t)$ とし，以下のようにモデル化した（図 (4.1(a))）．但し，本研究では，雑音のみを取り扱うため，インパルス応答の影響は無いものとしている．

$$y(t) = h(t) * x(t) + n(t) \quad (4.1)$$

$$h(t) = e_h(t)c_h(t) \quad (4.2)$$

$$x(t) = e_x(t)c_x(t) \quad (4.3)$$

$$n(t) = e_n(t)c_n(t) \quad (4.4)$$

$$\langle c_l(t), c_l(t - \tau) \rangle = \delta(t - \tau) \quad (4.5)$$

ここで， $e_x(t)$ ， $e_h(t)$ と $e_n(t)$ は， $x(t)$ ， $h(t)$ と $n(t)$ のエンベロープ， $c_x(t)$ ， $c_h(t)$ と $c_n(t)$ はキャリアでランダム変数， $\langle \cdot \rangle$ は集合平均である．

4.2 信号とパワーエンベロープの関係

前項でのモデルでは，下記のように2乗集合平均を取ることで，観測パワーエンベロープ $e_y^2(t)$ を得ることができる．

$$\langle y^2(t) \rangle = \langle h^2(t) * x^2(t) \rangle + \langle n^2(t) \rangle \quad (4.6)$$

ここで $\langle y^2(t) \rangle$ は，

$$\langle y^2(t) \rangle = \langle e_y^2(t)c_y^2(t) \rangle \quad (4.7)$$

$$= e_y^2(t) \langle c_y^2(t) \rangle \quad (4.8)$$

$$= e_y^2(t) \quad (4.9)$$

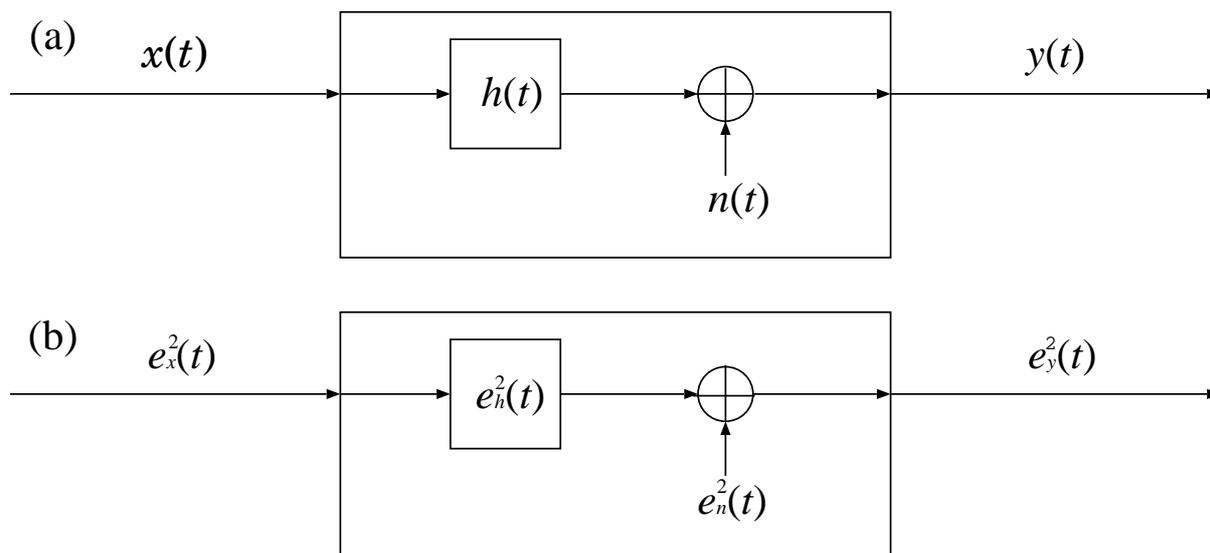


図 4.1: 伝達関数の構成 (a) は信号の場合. (b) はパワーエンベロープの場合 .

となっている . 但し , $\langle \mathbf{c}_y^2(t) \rangle = \delta(0) = 1$ である . $\langle \mathbf{n}^2(t) \rangle$ についても同様にして $e_n^2(t)$ を得ることができる . また , $\langle \mathbf{h}^2(t) * \mathbf{x}^2(t) \rangle$ については ,

$$\langle \mathbf{h}^2(t) * \mathbf{x}^2(t) \rangle = \left\langle \left\{ \int_{-\infty}^{\infty} \mathbf{h}(t - \tau) \mathbf{x}(\tau) d\tau \right\}^2 \right\rangle \quad (4.10)$$

$$= \left\langle \int_{-\infty}^{\infty} e_x(\tau_1) \mathbf{c}_x(\tau_1) e_h(t - \tau_1) \mathbf{c}_h(t - \tau_1) \int_{-\infty}^{\infty} e_x(\tau_2) \mathbf{c}_x(\tau_2) e_h(t - \tau_2) \mathbf{c}_h(t - \tau_2) d\tau_1 d\tau_2 \right\rangle \quad (4.11)$$

$$= \int_{-\infty}^{\infty} e_x(\tau_1) e_h(t - \tau_1) \int_{-\infty}^{\infty} e_x(\tau_2) e_h(t - \tau_2) \langle \mathbf{c}_x(\tau_1) \mathbf{c}_x(\tau_2) \rangle \langle \mathbf{c}_h(t - \tau_1) \mathbf{c}_h(t - \tau_2) \rangle d\tau_1 d\tau_2 \quad (4.12)$$

ここで , $\tau_1 = \tau_2 = \tau$ とすると , 式 (4.5) との関係より

$$\langle \mathbf{h}^2(t) * \mathbf{x}^2(t) \rangle = \int_{-\infty}^{\infty} e_x(\tau_1) e_h(t - \tau_1) \int_{-\infty}^{\infty} e_x(\tau_2) e_h(t - \tau_2) \delta(\tau_2 - \tau_1)^2 d\tau_1 d\tau_2 \quad (4.13)$$

$$= \int_{-\infty}^{\infty} e_x^2(\tau) e_h^2(t - \tau) d\tau \quad (4.14)$$

$$= e_x^2(t) * e_h^2(t) \quad (4.15)$$

となる . これにより式 (4.6) は ,

$$e_y^2(t) = e_h^2(t) * e_x^2(t) + e_n^2(t) \quad (4.16)$$

となり，式(4.1)の信号の関係と式(4.16)パワーエンベロップの関係が等しくなる．本研究では，このパワーエンベロップの関係(図4.1(b))を用いる．但し，雑音環境を想定しているため，残響 $e_h^2(t)$ の影響は無いものとしている．

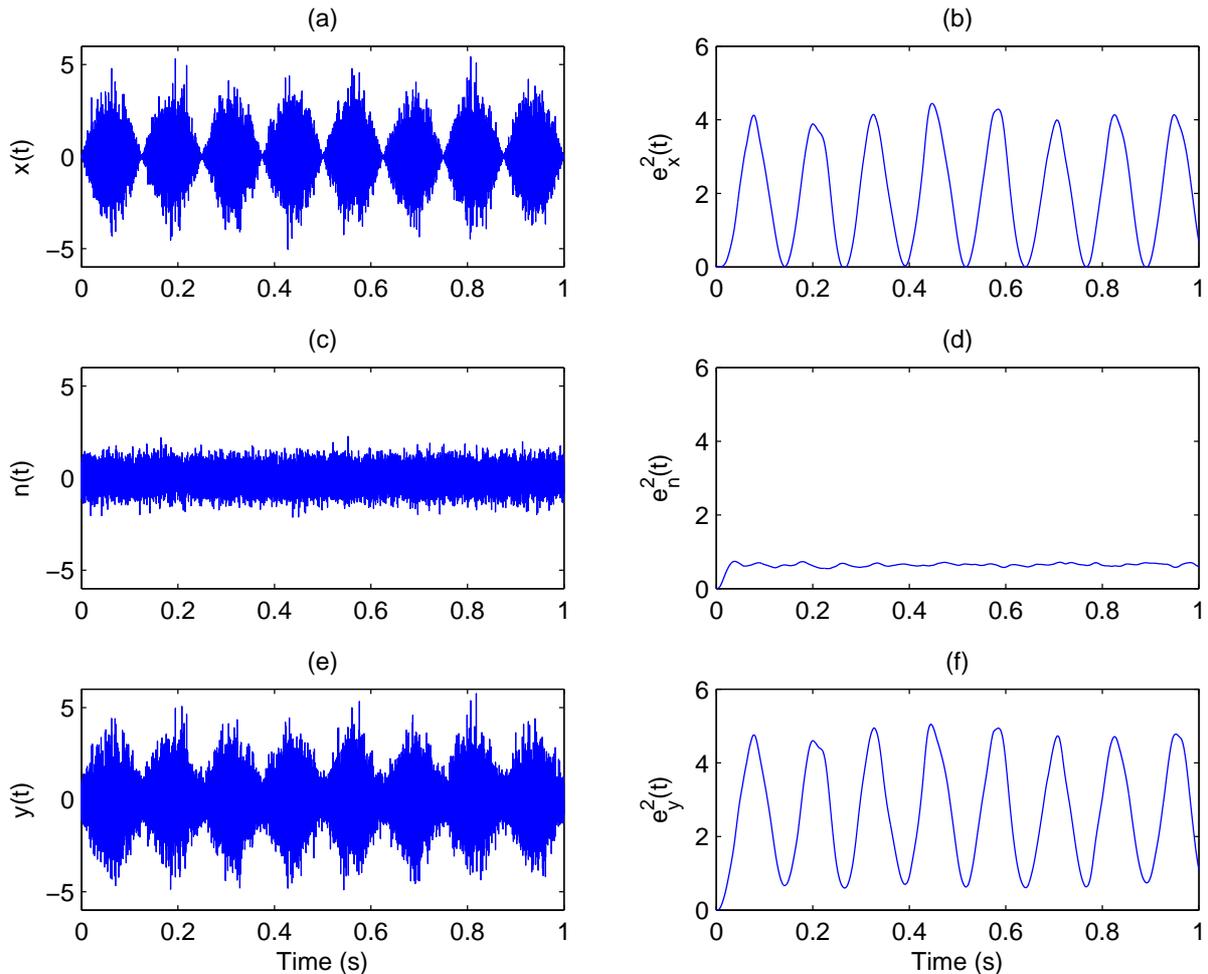


図 4.2: 信号とパワーエンベロップの関係

図 4.2 は式 (4.1) と式 (4.16) の信号とパワーエンベロップの関係を時間波形で示した例である．但し，残響の影響はないものとしている．図 4.2(a) は入力信号，(c) は雑音信号，(e) は観測信号である．入力信号に雑音信号が加算されて観測信号になっている．また (b) は入力パワーエンベロップ，(d) は雑音パワーエンベロップ，(f) は観測パワーエンベロップである．信号と同様に，入力パワーエンベロップに雑音パワーエンベロップが加算されて，観測パワーエンベロップになっているのが分かる．

4.3 パワーエンベロープの抽出法

MTFに基づく雑音抑圧法を提案するには，観測信号 $y(t)$ から，観測パワーエンベロープである $e_y^2(t)$ を抽出する必要がある．本研究では，以下の方法で出力信号 $y(t)$ から，パワーエンベロープを抽出する．

$$\hat{e}_y^2(t) = \text{LPF} \left[|y(t) + j\text{Hilbert}\{y(t)\}|^2 \right] \quad (4.17)$$

ここで $\text{LPF}[\cdot]$ はローパスフィルタ， $\text{Hilbert}(\cdot)$ はヒルベルト変換である．この方法は信号の瞬時振幅の計算に基づいている．また，パワーエンベロープの高周波成分を取り除くために，後処理としてローパスフィルタリングを行なう．ローパスフィルタのカットオフ周波数は 20 Hz とした．この手法は Unoki ら [31] によって有効性が示されている．またカットオフ周波数は，Arai ら [35] や金寺ら [36][37] によって，音声知覚と音声認識では，1～16 Hz の変調周波数が重要であるという報告に基づいて設定した．

4.4 MTF に基づくパワーエンベロープ逆フィルタ法

広林 [29]，Unoki[31]-[33] らが行なったパワーエンベロープ逆フィルタ法について説明する．彼らは残響環境を想定しているため，式 (4.16) は，

$$e_y^2(t) = e_h^2(t) * e_x^2(t) \quad (4.18)$$

となる．式 (3.8) から， $e_h(t)$ を $e_h^2(t) = a \exp\left(-\frac{6.9t}{T_R}\right)$ とする．式 (4.16) も式 (4.18) も実際には，離散時間で利用するため，周波数変換として， z 変換を用いる．連続時間 t を離散時間系列 k とし， z 変換を $Z[\cdot]$ とすると， $e_h^2(t)$ は z 領域で

$$\mathbf{Z}[e_h^2(k)] = \frac{a^2}{1 - \exp\left(-\frac{13.8}{T_R \cdot f_s}\right) z^{-1}} \quad (4.19)$$

と表現される [31]．ここで， f_s はサンプリング周波数である．また，式 (4.18) の関係が $\mathbf{Z}[e_y^2(k)]/\mathbf{Z}[e_x^2(k)] = \mathbf{Z}[e_h^2(k)]$ となることから，入力パワーエンベロープの変調スペクトルは，

$$\mathbf{Z}[e_x^2(k)] = \frac{\mathbf{Z}[e_y^2(k)]}{a^2} \left\{ 1 - \exp\left(-\frac{13.8}{T_R \cdot f_s}\right) \right\} z^{-1} \quad (4.20)$$

で求めることができる．このように，パワーエンベロープ逆フィルタ法はインパルス応答の逆特性を用いて，入力パワーエンベロープを得る手法である．ここで，残響環境での MTF の算出式である式 (3.7) より，残響環境での MTF は，インパルス応答をフーリエ変換し，その直流分で正規化することであり，伝達関数と考えることができる．パワーエンベロープ逆フィルタ法は， z 領域ではあるが，観測パワーエンベロープの変調スペクトルを伝達関数で割っている．つまりは，MTF， $1/m(f_m)$ を観測変調スペクトルに掛けることで MTF を回復していることと等しい．

第5章 提案法

5.1 雑音・残響抑圧の処理体系

本研究の目標である雑音・残響の同時抑圧処理を行なうには，4章で述べたパワーエンベロープ逆フィルタ法が $1/m(f_m)$ を掛けることで，残響抑圧を行なっていることから，3章で述べた雑音・残響環境での MTF(式(3.10)) の逆数を掛けることができれば，入力パワーエンベロープを推定し，雑音・残響抑圧ができる．ここで，残響環境では，Unoki らが既に MTF に基づいたパワーエンベロープ逆フィルタ法を実現しているが，雑音環境では，抑圧法は提案されていない．雑音・残響抑圧法の実現するには，まず MTF に基づいた雑音抑圧法の検討がされるべきである．そこで本研究で，Unoki らと同様に $1/m(f_m)$ 倍することで雑音抑圧を行なう手法を提案する．

5.2 雑音抑圧法を提案する際の問題点

MTF に基づく雑音抑圧法を提案する際に，以下の問題点が挙げられる．

1. 雑音により影響を受けているのは，パワーエンベロープの変調度だけではない．
2. MTF を算出するのに入力パワーエンベロープの情報が必要である．

1つめの問題について，雑音により，入力パワーエンベロープの変調度だけではなく，振幅も影響を受けているため，ただ単に変調度を回復するために，MTF の逆数である $m(f_m)$ を観測パワーエンベロープに掛けるだけでは，入力パワーエンベロープが得られない．2つめの問題について，雑音環境下の MTF の算出には式(3.6) より，入力の情報である $\overline{e_x^2}$ が必要である．そのため，雑音環境での MTF を推定しなければならない．この2つの問題点を解決すれば，MTF に基づく雑音抑圧法を提案できる．

5.3 提案法のコンセプト

提案法のコンセプトは，雑音の影響を受けて低下した MTF を回復することで，観測パワーエンベロープ $e_y^2(t)$ から入力パワーエンベロープ $e_x^2(t)$ を得ることである．具体的には $1/m(f_m)$ を掛けることで，MTF を回復するが，観測パワーエンベロープ $e_y^2(t)$ は，雑音に

より変調度だけではなく、振幅も影響を受けている。そのため、変調度の回復を行なう前に、振幅の回復を行なう。

5.4 MTF に基づいた雑音抑圧法

5.4.1 MTF に基づいた雑音抑圧法の原理

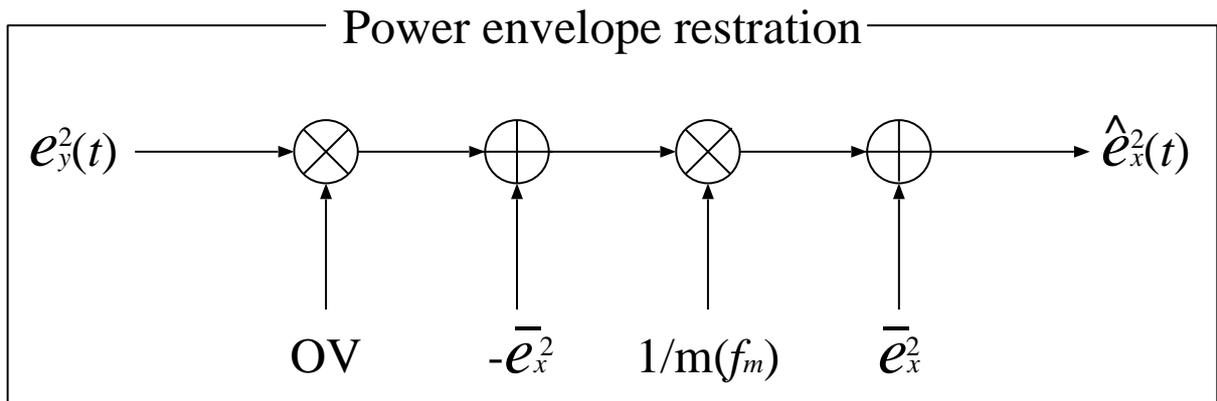


図 5.1: MTF に基づいた雑音抑圧法の構成

MTF に基づいた雑音抑圧法は図 (5.1) に示す手順で行なわれる。観測パワーエンベロープ $e_y^2(t)$ は、雑音の影響により、振幅と変調度に影響を受けていることが式 (3.5) より分かる。そこで、振幅と変調度の回復を行なうことが、MTF に基づいた雑音抑圧法となる。まず振幅の補正を行なう。

$$\mathbf{OV} = \frac{\overline{e_x^2}}{\overline{e_x^2} + \overline{e_n^2}} \quad (5.1)$$

式 (5.1) は、振幅回復のための補正值である。この式 (5.1) を式 (3.5) にかける。これにより、次のような式を得る。

$$\overline{e_x^2} + \overline{e_x^2} \cdot m(f_m) \cdot \cos(2\pi f_m t) \quad (5.2)$$

式 (5.2) は、式 (3.5) から振幅が回復された状態になっている。この状態から低下した変調度を、 $1/m(f_m)$ を掛けることで回復するが、式 (5.2) に対して、そのまま $1/m(f_m)$ を掛けると回復した振幅が再び影響を受けてしまう。そのため、式 (5.2) の第 2 項に対して、 $1/m(f_m)$ を掛けなければならない。 $\overline{e_x^2}$ は、式 (5.2) 時のパワーエンベロープの時間平均値であるため、容易に得ることができる。式 (5.2) から、 $\overline{e_x^2}$ を差し引くと

$$\overline{e_x^2} \cdot m(f_m) \cdot \cos(2\pi f_m t) \quad (5.3)$$

となる．式(5.3)の状態にして， $1/m(f_m)$ を掛けて，変調度の回復が行なえる．

$$\overline{e_x^2} \cdot m(f_m) \cdot \cos(2\pi f_m t) \times \frac{1}{m(f_m)} = \overline{e_x^2} \cdot \cos(2\pi f_m t) \times \frac{1}{m(f_m)} \quad (5.4)$$

そして，差し引いていた $\overline{e_x^2}$ を加算すると

$$\overline{e_x^2} \cdot \cos(2\pi f_m t) \times \frac{1}{m(f_m)} + \overline{e_x^2} = \overline{e_x^2}(1 + \cos(2\pi f_m t)) \quad (5.5)$$

となる，結果として入力パワーエンベロープ $e_x^2(t)$ を得ることができる．最後に負のパワーが無いという仮定から，負の値になっている部分を0にする．この過程がMTFに基づく雑音抑圧法である．図(5.2)の(a)から(e)は，観測パワーエンベロープと式(5.2)から式(5.5)までのパワーエンベロープ(実線)である．(f)は観測パワーエンベロープに対して，単に $1/m(f_m)$ を掛けた場合のパワーエンベロープである．破線は入力パワーエンベロープを表している．単に $1/m(f_m)$ 倍を行なっただけでは，入力パワーエンベロープを得ることができないのが，図(5.2(f))から見て取れる．また，図(5.2(e))より，提案法が入力パワーエンベロープを精度良く推定できていることが分かる．

5.4.2 雑音環境でのMTFの推定

式(3.6)より，MTFの算出には， $\overline{e_x^2}$ と $\overline{e_n^2}$ が必要になる． $\overline{e_n^2}$ は無音声区間から推定を行なう． $\overline{e_x^2}$ は推定した $\overline{e_n^2}$ を $\overline{e_y^2}$ から差し引くことで得る．これにより雑音環境でのMTFを算出する．

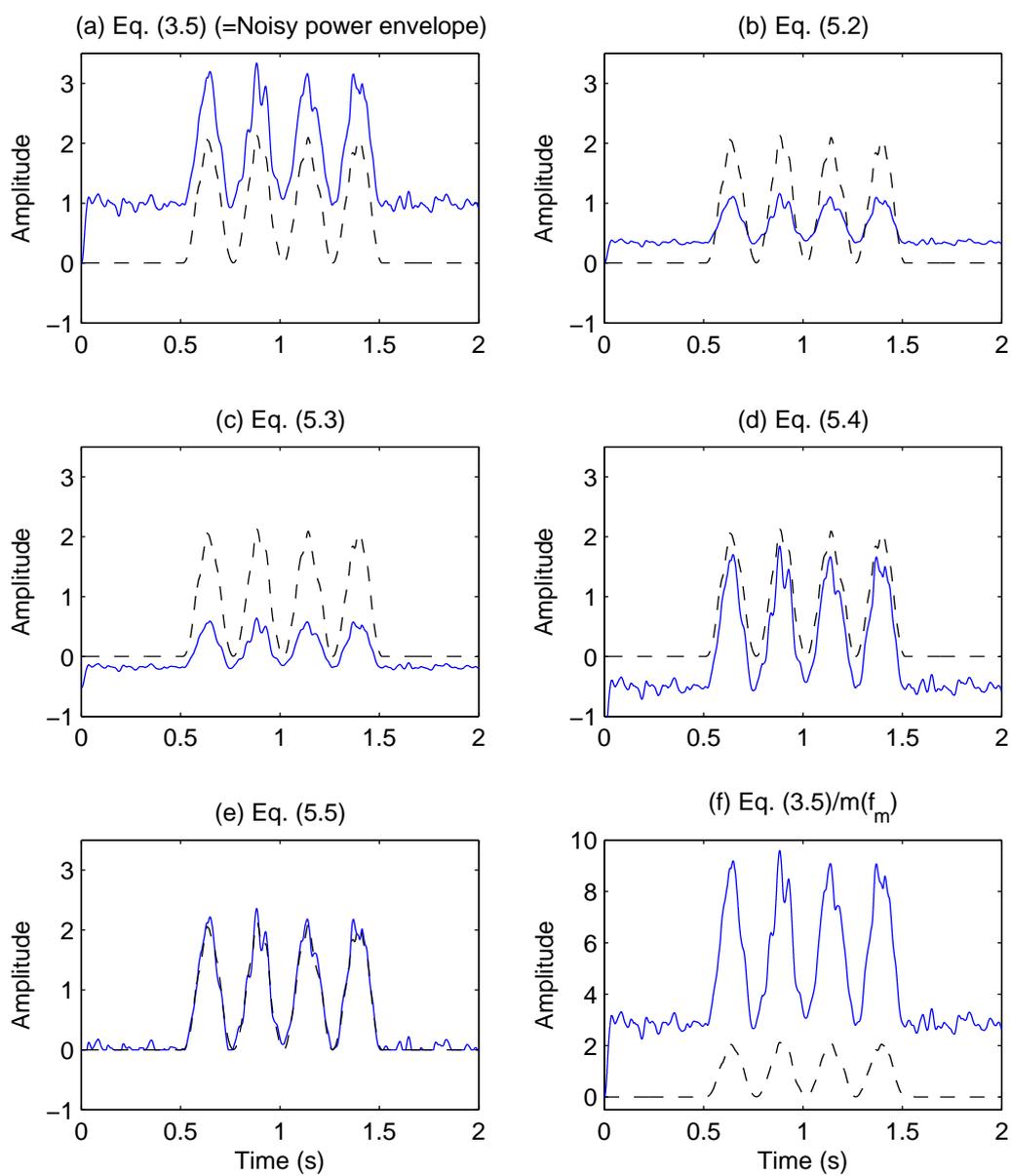


図 5.2: パワーエンベロープでのMTFに基づいた雑音抑圧法の流れ . (a) から (e) の実線は各式でのパワーエンベロープの状態 , (f) は観測パワーエンベロープに $1/m(f_m)$ 倍した場合 . 破線は , 入力パワーエンベロープ .

5.5 音声信号への適応と帯域分割処理

音声信号のパワーエンベロープは、全帯域において共変調してはいない。音声信号に対してMTFに基づく雑音抑圧処理を行なう時には、音声のパワーエンベロープが共変調でない場合を考慮しなければならない。そこで帯域分割処理を行なうことで、共変調とみなせる帯域毎に分割し、各帯域毎にMTFに基づく雑音抑圧法を施す必要がある。帯域分割幅については、キャリアの無相関性とMTFのパワーエンベロープの関係がトレードオフであるため、両者の条件を満たす帯域幅を決定しなければならない。条件を満たす帯域幅については、Unokiら [31] によって検討されており、1チャンネルあたりの帯域幅は100 Hzが良いとしている。これに基づき提案法も帯域幅を100 Hzとした。図(5.3)は、提案法をSNR 0 dBの雑音音声(/aikawarazu/)に施した結果の信号波形である。雑音が抑圧されているのが見て取れる。図(5.4)は提案法のブロック図である。提案法を行なう手順は

1. 観測音声を、定帯域フィルタバンクで1チャンネルあたり100 Hz毎に帯域分割処理を行なう。
2. 帯域分割した観測音声からパワーエンベロープ $e_y^2(t)$ を抽出する。
3. パラメータを推定する。また、抽出したパワーエンベロープを用いてキャリアを抽出する。
4. 各パワーエンベロープに対して回復処理を行ない、パワーエンベロープを回復する。
5. キャリアと掛け合わせて、合成処理を行うことで推定入力信号を得る。

となっている。

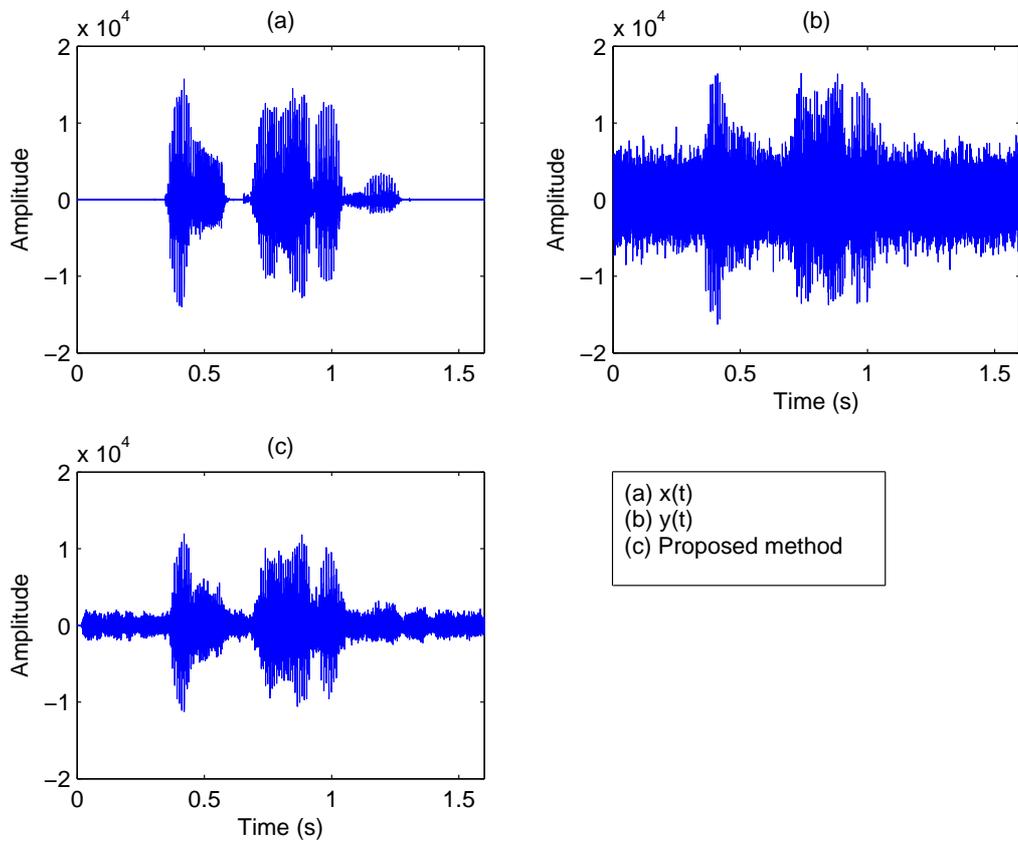


図 5.3: 入力音声 $x(t)$, 観測音声 $y(t)$ と提案法を施した音声 (SNR 0 dB, 音声は/aikawarazu/)

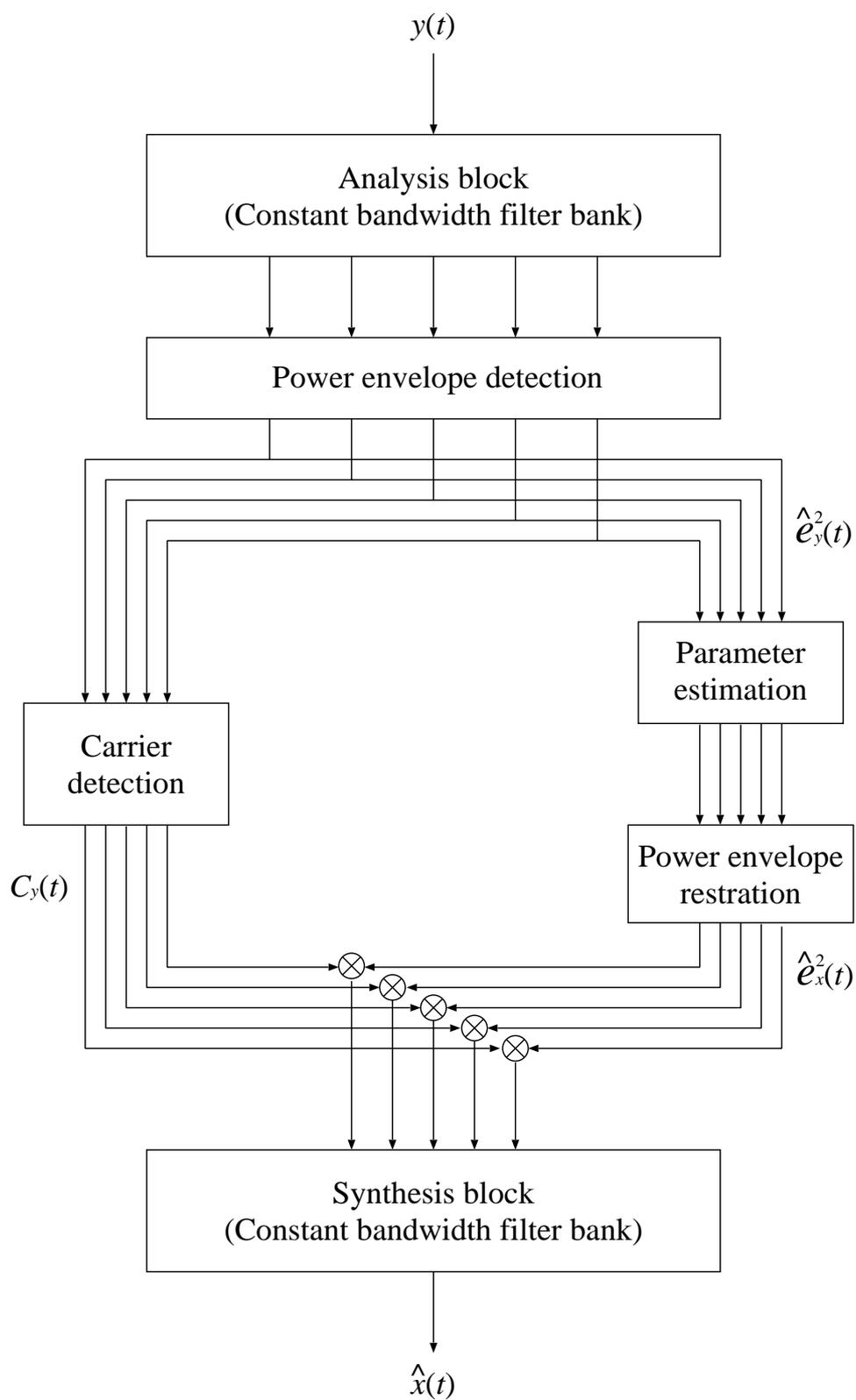


図 5.4: 提案法のブロック図

5.6 パワーエンベロープを差し引く手法

提案法は雑音パワーエンベロープの平均値を観測パワーエンベロープから差し引く手法と等価である．これは，式(4.16)と雑音パワーエンベロープが時間に対して一定という仮定より

$$\begin{aligned} e_y^2(t) &= e_x^2(t) + e_n^2(t) \\ &= e_x^2(t) + \overline{e_n^2(t)} \end{aligned} \quad (5.6)$$

観測パワーエンベロープを上記の式で表すことができる，この式(5.6)から雑音パワーエンベロープの平均値を差し引けば，

$$e_y^2(t) - \overline{e_n^2(t)} = e_x^2(t) \quad (5.7)$$

となり，入力パワーエンベロープを得る．そのため結果としては，提案法と等価となる．パワーエンベロープを差し引く手法は，雑音パワーエンベロープを差し引くことしかできず，残響の影響を抑えることは不可能である．しかし，提案法はMTFに基づいているため，同じ概念に基づいているUnokiらの手法と組み合わせる事で残響の影響を抑えられる可能性がある．

第6章 提案法の評価

この章では，提案法が雑音に対して効果があるか調べるために，評価シミュレーションを行なう．パワーエンベロープの相関値とパワー比の改善度を用いて，観測パワーエンベロープ $e_y^2(t)$ からパワーエンベロープが回復できているか評価する．また雑音音声から，どれだけ音声を回復できたかを調べるために，対数スペクトル距離を用いて評価する．

6.1 条件

評価に用いる音声は，ATR データベース [38] にある男性 5 名 (mau, mht, mnm, mtm, mtt)，女性 5 名 (faf, ffs, fkn, fsu, fyn)，計 10 名の話者が発話した単語 (/aikawarazu/, /shinbun/, /joudan/) とした．SNR が 20, 10, 5, 0, -5 dB になるように白色雑音を付加した．1 つの SNR に対して白色，ピンク，バブル雑音をそれぞれ 100 個用意した．そのため雑音を加算された出力信号である $y(t)$ の総数は， $10 \times 3 \times 5 \times 100 = 45000$ となった．これらの音声に，1 チャンネルあたり 100 Hz で帯域分割処理をおこない，提案法を帯域ごとに施した．データベースのサンプリング周波数が 2 kHz であるため，総チャンネル数は 100 チャンネルである．MTF の推定は，無音声区間と音声区間の切り分けが理想的にできたものとして行った．また従来の雑音抑圧法との効果を比較するために，ウィナーフィルタリング [6]，SS 法 [4]，MMSE-STSA 法 [10] を用いた．

評価項目として，各チャンネルにおける相関値 (Corr)，パワー比 (SNR) の改善度，対数スペクトル距離 (LSD) と音声明瞭度と関係が取られている重み付けをした LSD [39] を用いた．相関値，パワー比と LSD は下記式で表される．

$$\text{Corr}(e_x^2, \hat{e}_x^2) = \frac{\int_0^T (e_x^2(t) - \overline{e_x^2(t)}) (\hat{e}_x^2(t) - \overline{\hat{e}_x^2(t)}) dt}{\sqrt{\left\{ \int_0^T (e_x^2(t) - \overline{e_x^2(t)}) dt \right\} \left\{ \int_0^T (\hat{e}_x^2(t) - \overline{\hat{e}_x^2(t)}) dt \right\}}} \quad (6.1)$$

$$\text{SNR}(e_x^2, \hat{e}_x^2) = 10 \log_{10} \frac{\int_0^T (e_x^2(t))^2 dt}{\int_0^T (e_x^2(t) - \hat{e}_x^2(t))^2 dt} \quad (6.2)$$

$$\text{LSD} = \sqrt{\frac{1}{W} \sum_{\omega} \left(20 \log_{10} \frac{|S_x(\omega)|}{|\hat{S}_x(\omega)|} \right)^2} \quad (6.3)$$

ここで， $\overline{e_x^2}$ は $e_x^2(t)$ の平均， $e_x^2(t)$ と $\hat{e}_x^2(t)$ は入力パワーエンベロープと回復されたパワーエンベロープ， W は周波数の上限 (10 kHz)， $S_x(\omega)$ には入力信号の振幅スペクトル， $\hat{S}_x(\omega)$

は、観測信号，提案法を施した信号の振幅スペクトルである．パワーエンベロープの相関とパワー比を調べることで，パワーエンベロープがどのくらい復元できたか，LSDの改善度を調べることで，どのくらい雑音音声を回復したか，重み付きLSDの改善度を調べることで，音声明瞭度がどのくらい改善されたか，客観的に評価できる．

6.2 結果と考察

6.2.1 相関，SNRの改善度による評価

各図のバーが相関値の改善度とパワー比の改善度の平均値を，エラーバーがその標準偏差を示している．横軸は帯域分割した際のチャンネルナンバー，縦軸は相関値とパワー比の改善度である．

提案法について，図 6.1 は白色雑音，図 6.2 はピンク雑音，図 6.3 はバブル雑音による結果である．相関値は雑音の種類に関係なく，おおむね ± 0.15 に分布している，これは回復後のパワーエンベロープが，回復前のパワーエンベロープから雑音パワーエンベロープの平均パワー分を差し引いたものであるため，パワーエンベロープの形に，あまり影響をあたえないためである．パワー比の改善度について，白色雑音の場合，SNRが低くなるにつれて，パワー比の改善度が増加している．低帯域側のチャンネルにおいて，改善度が低いのは，音声の主要な成分が低域にあるため，低帯域側でのSNRが元々高いため，改善度として少量になっているためである．ピンク雑音の場合も白色雑音と同様の傾向を示している．バブル雑音の場合，白色，ピンク雑音の場合よりもバブル雑音の改善度は低いものとなっている．バブル雑音は，音声の足し合わせから成る雑音であるため，時間軸上での雑音パワーエンベロープの分散が大きい．これにより提案法の効果が，白色，ピンク雑音よりも低くなっている．しかし，SNRが低くなるごとに改善度が増加する傾向は見られ，効果はある．

これらの結果より，提案法は雑音が付加された観測パワーエンベロープからパワーエンベロープを復元する点において効果があることが言える．

従来法との比較について，白色雑音，ピンク雑音の場合，SNR -5 dBの時に，パワー比の改善度において，提案法は従来法より劣る傾向が見られるが，SNR 20 dB時のように，改善度が大幅にマイナスになることはない．相関値の改善度に関しては，提案法も従来法も同程度である．バブル雑音の場合は，SNR 0 dBとSNRが低い時でも従来法の改善度と比べ，パワー比の改善度が劣ることはなく，SNR 20 dBでは，改善度の差は顕著に現れている．また相関値の改善度に関して，従来法では，高めの周波数帯域では改悪の傾向が見られるが，提案法では，改悪の傾向は見られない．

これらのことから，提案法は，雑音付加時より，パワー比が改悪にならない分，他の手法よりパワーエンベロープを回復する点で優れている．

6.2.2 LSD による評価

図 6.13 は白色雑音，図 6.14 はピンク雑音，図 6.15 はバブル雑音による結果である．各図のバーが LSD の改善度の平均値をエラーバーがその標準偏差を示している．

提案法について，3 種類の雑音とも SNR が低くなるにつれて改善度が増加している．白色雑音の場合，最大で約 31 dB の改善度が見られる．ピンク雑音の場合は約 28 dB の改善度，バブル雑音の場合は約 11 dB の改善度となっている．バブル雑音，SNR が 20 dB の場合に改悪の傾向が見られる．これは SNR 20 dB と雑音成分が音声成分に比べて少ないため雑音パワーエンベロープの引き過ぎによるものと考えられる．

このことから雑音信号から音声信号を回復する点で，提案法は効果があることが示された．

従来法との比較について，SS 法と MMSE-STSA 法と比較すると，バブル雑音の SNR - 5 dB の時に約 2 dB の差があるが，他の SNR，雑音の場合では，ほぼ同じ改善度であることが確認できる．ウィナーフィルタリングと比較する．白色，ピンク雑音では，SNR が低くなるほどに，提案法の改善度よりも従来法の改善度が大きくなっているため，ウィナーフィルタリングによる手法の方が優れている．しかし，バブル雑音では改善度の差は，ないと言ってもよいほどになくなっており，バブル雑音に関しては同程度の効果と考えられる．

このことより，LSD において，提案法はウィナーフィルタリングには劣る傾向があるが，SS 法，MMSE-STSA 法と同程度の効果があることが示された．

6.2.3 重み付け LSD による評価

図 6.16 は白色雑音，図 6.17 はピンク雑音，図 6.18 はバブル雑音による結果である．各図のバーが重み付け LSD の改善度平均値をエラーバーがその標準偏差を示している．

提案法について，3 種類の雑音とも SNR が低くなるにつれて改善度が増加している．白色雑音の場合，最大で約 8 dB の改善度が見られる．ピンク雑音の場合は約 8 dB の改善度，バブル雑音の場合は約 5 dB の改善度となっている．

従来法との比較について，ピンク雑音とバブル雑音の場合，提案法は従来法と同程度の改善度であることが分かる．白色雑音の場合，SNR が低くなるにつれて，ウィナーフィルタリングに対して最大で約 3 dB ほど劣る傾向がみられる．しかし，LSD の改善度での評価よりも提案法とウィナーフィルタリングの差は少ない．

このことにより，重み付き LSD において，提案法は白色雑音の場合，ウィナーフィルタリングには劣る傾向があるが，SS 法，MMSE-STSA 法と同程度の効果があること，ピンク雑音とバブル雑音の場合では，従来法と同程度の効果があることが示された．

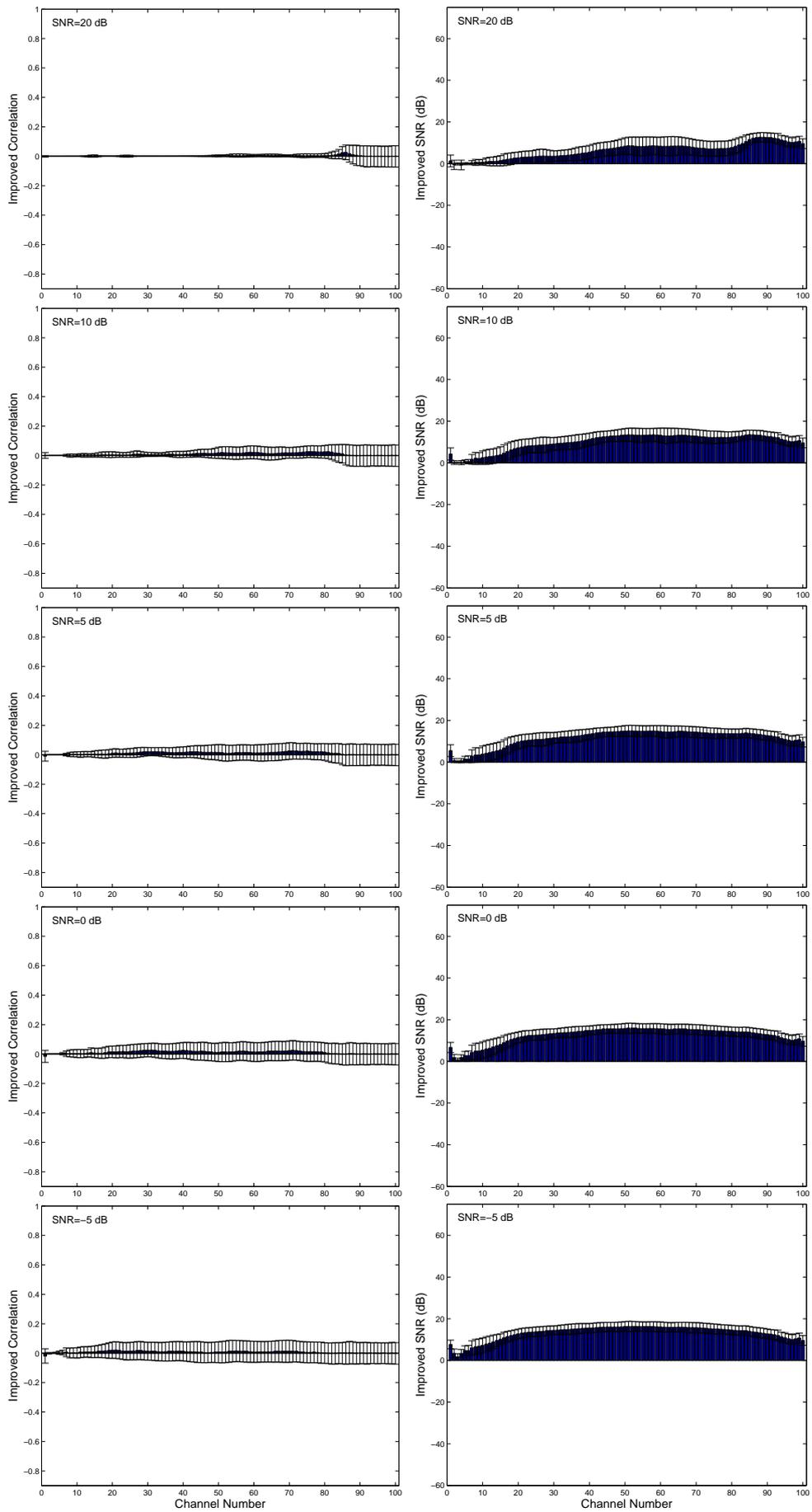


図 6.1: 提案法:白色雑音での相関値とパワー比の改善度

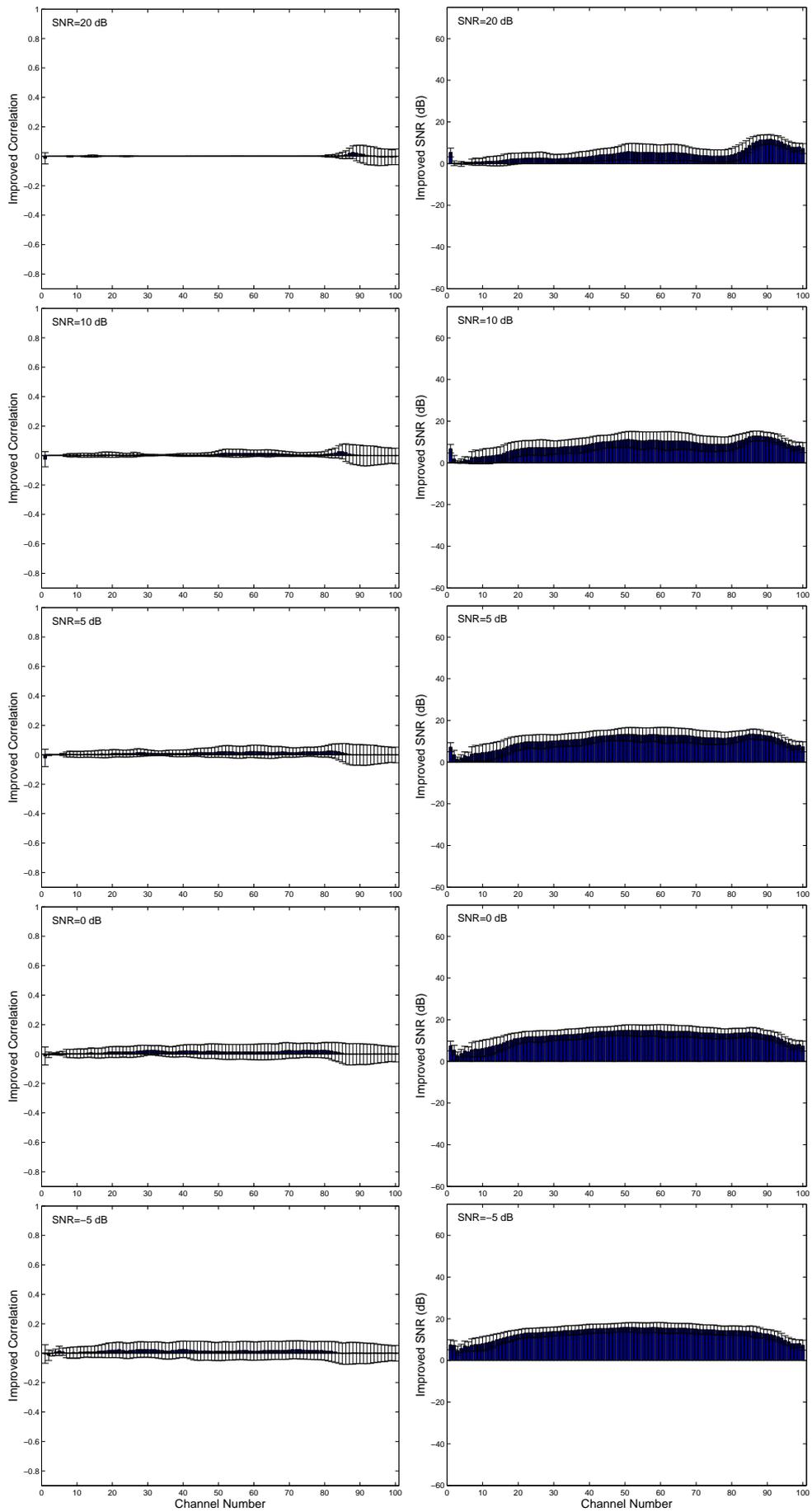


図 6.2: 提案法:ピンク雑音での相関値とパワー比の改善度

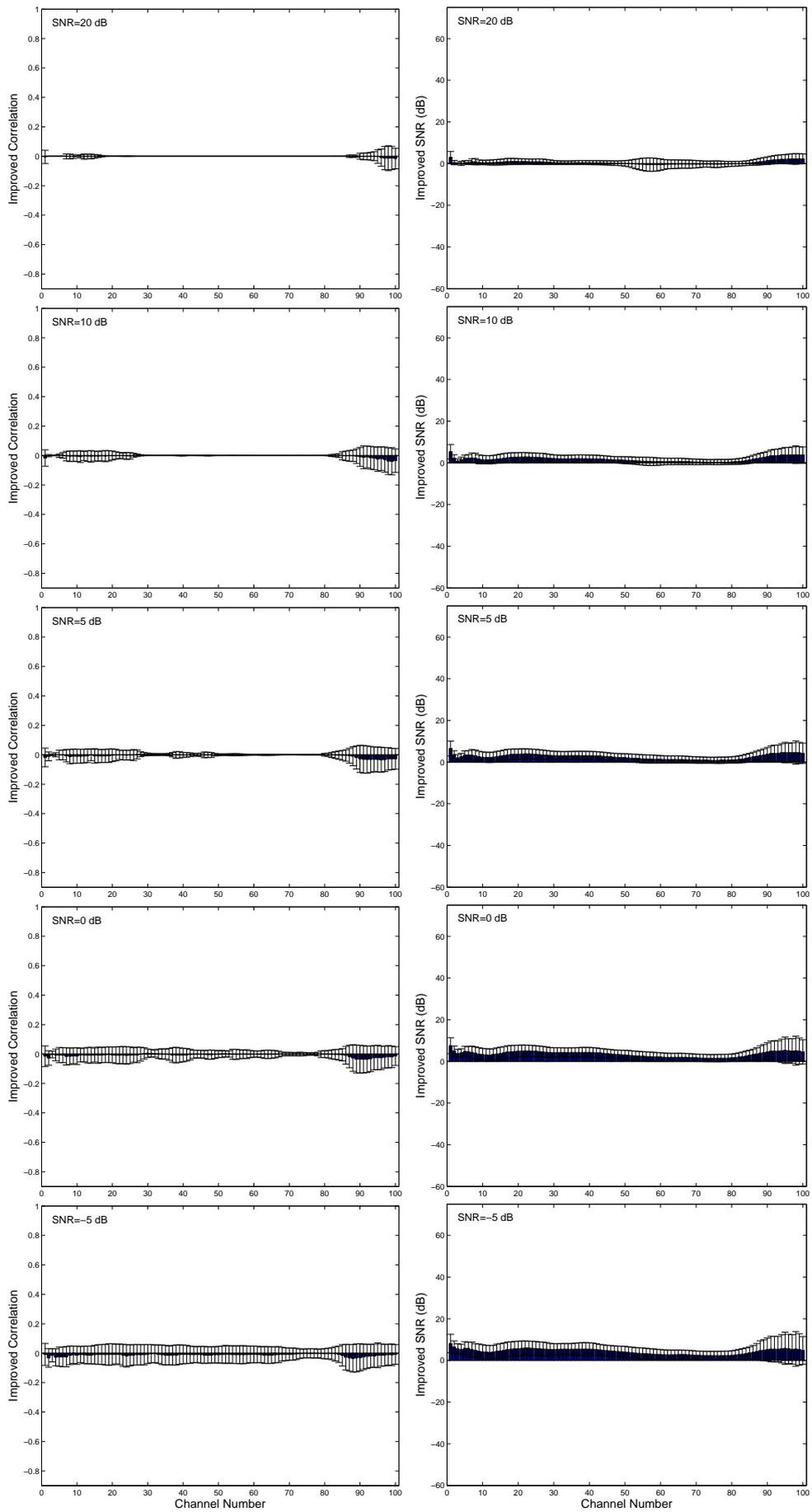


図 6.3: 提案法:バブル雑音での相関値とパワー比の改善度

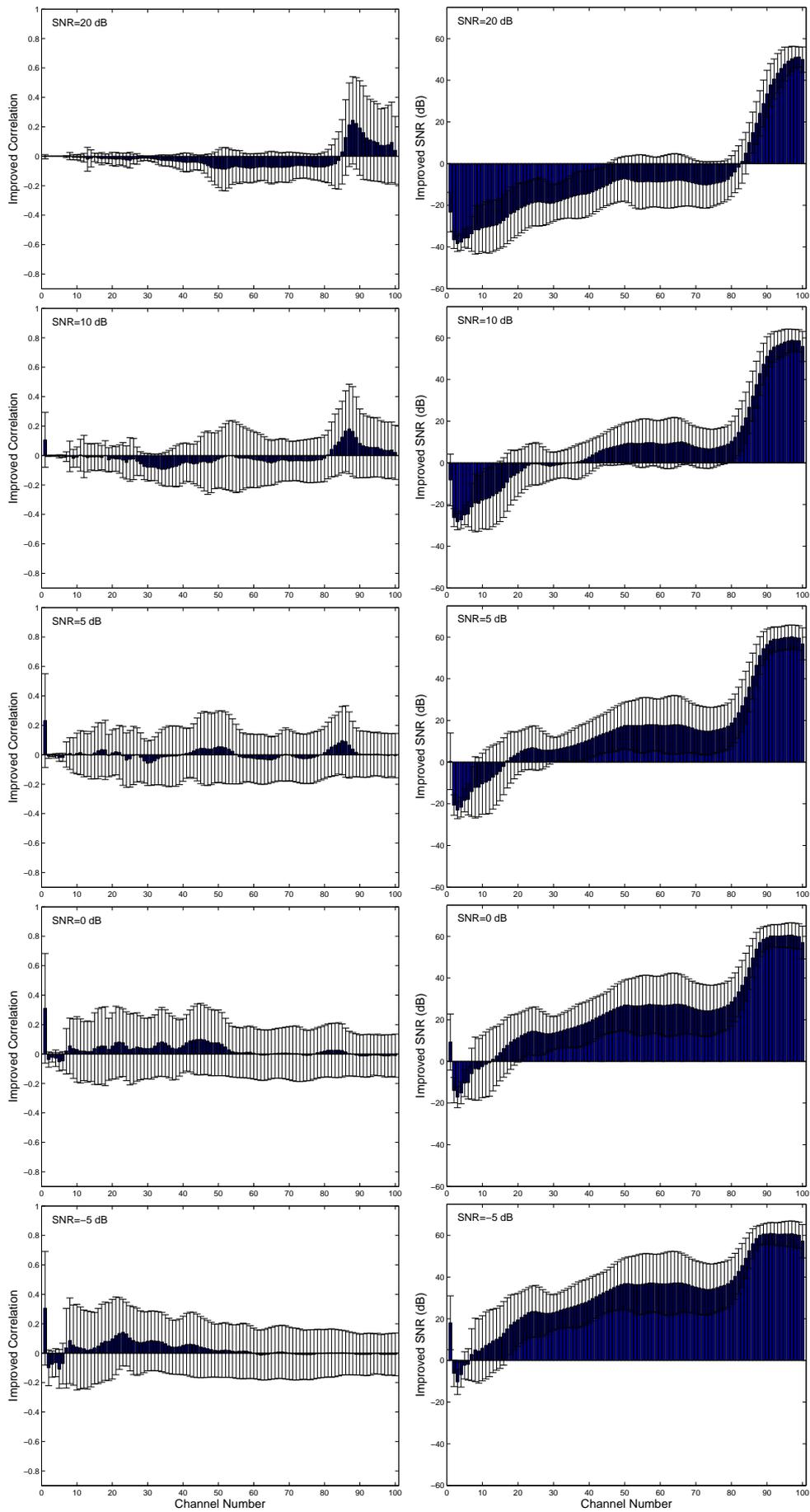


図 6.4: ウィナーフィルタ:白色雑音での相関値とパワー比の改善度

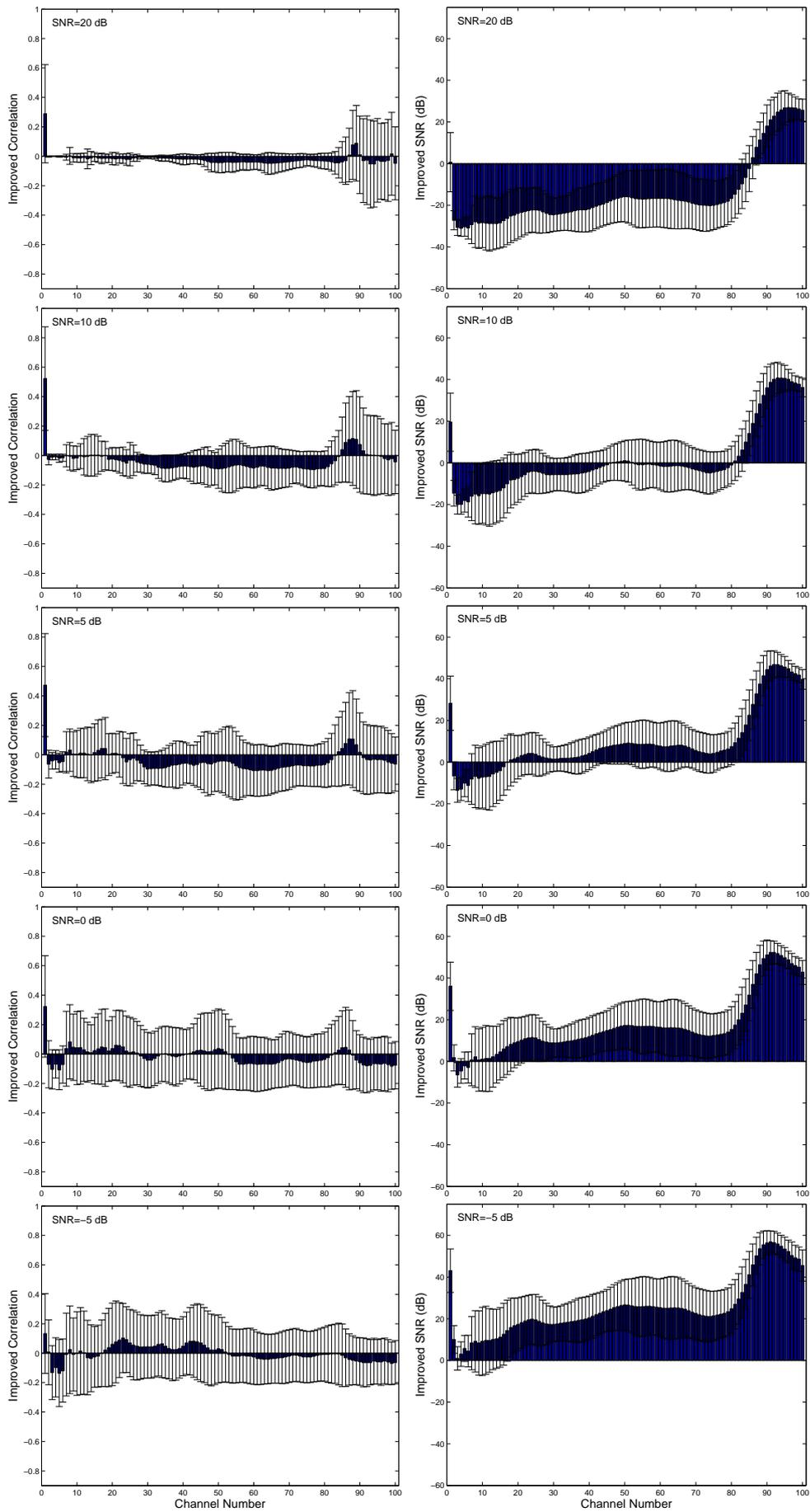


図 6.5: ウィナーフィルタ:ピンク雑音での相関値とパワー比の改善度

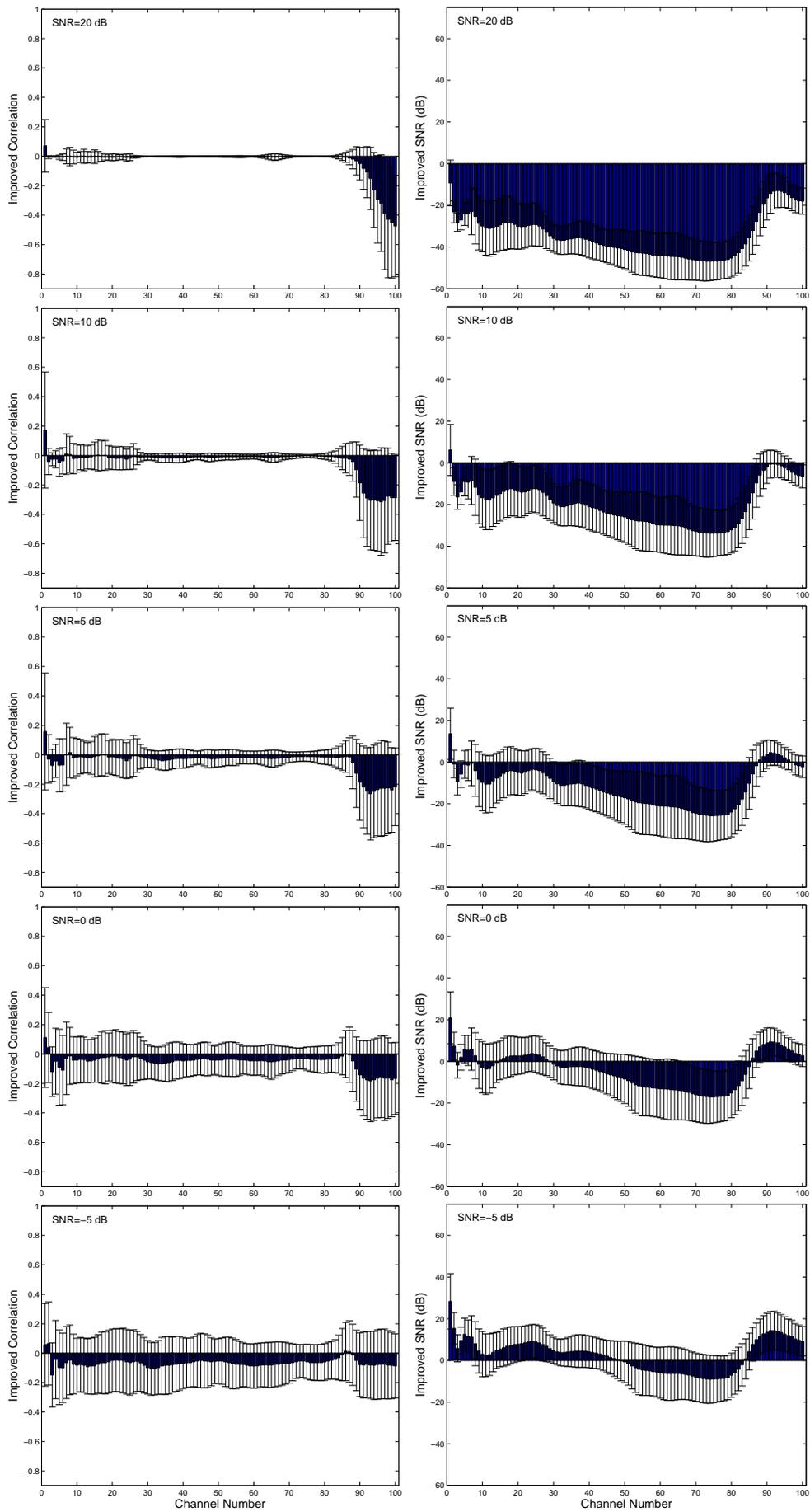


図 6.6: ウィナーフィルタ:バブル雑音での相関値とパワー比の改善度

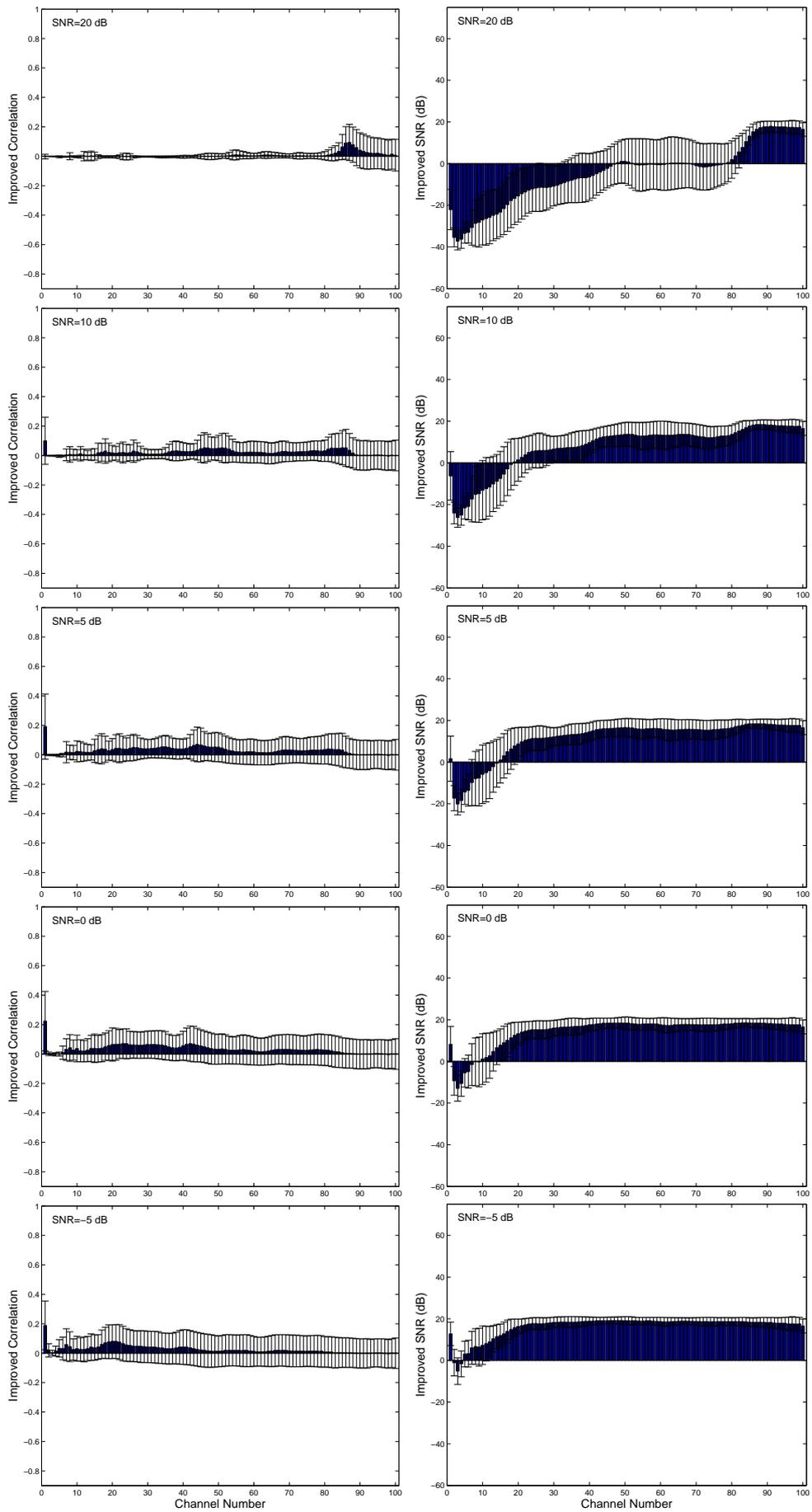


図 6.7: SS 法:白色雑音での相関値とパワー比の改善度

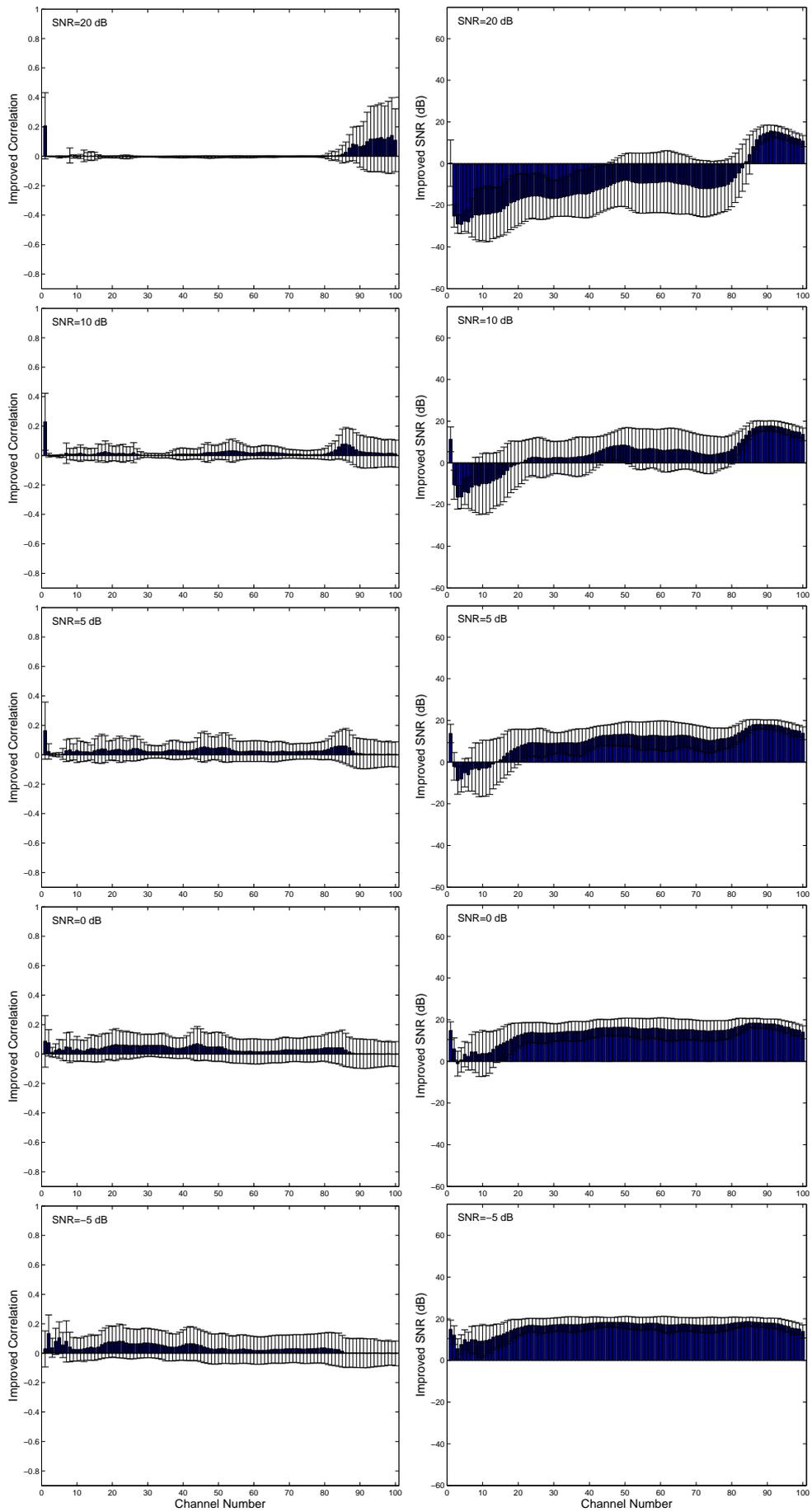


図 6.8: SS 法: ピンク雑音での相関値とパワー比の改善度

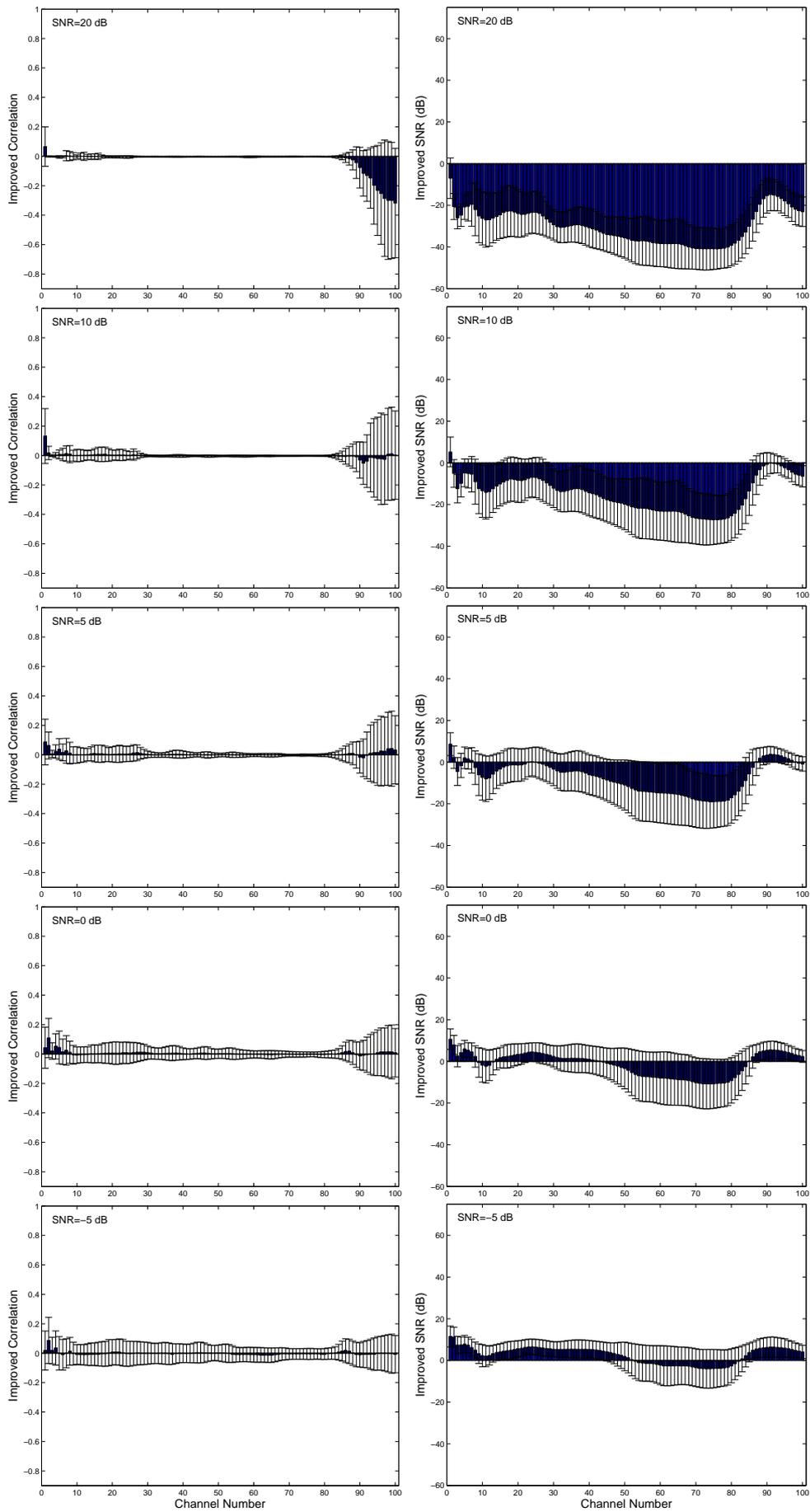


図 6.9: SS 法:バブル雑音での相関値とパワー比の改善度

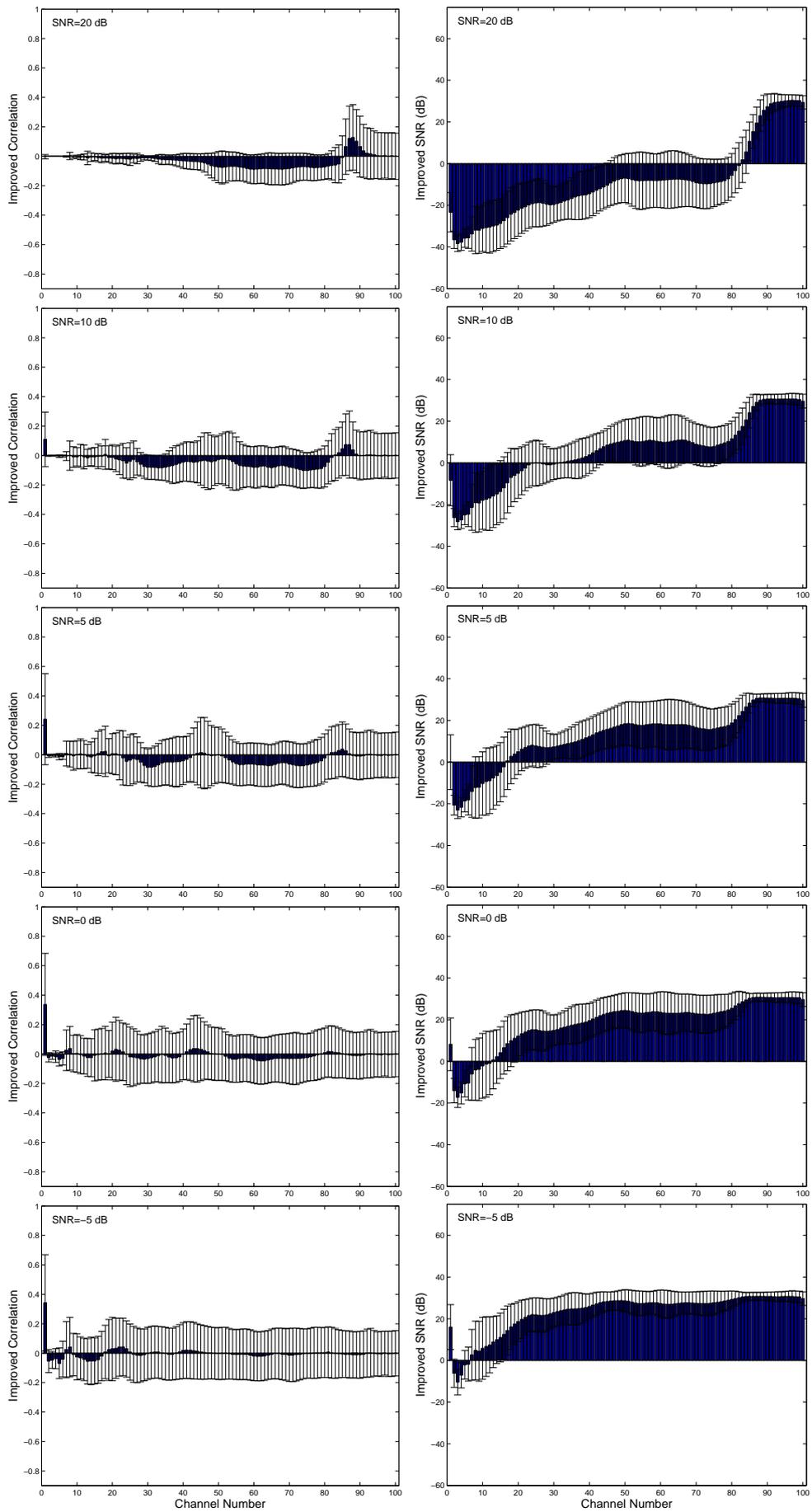


図 6.10: MMSE-STSA:白色雑音での相関値とパワー比の改善度

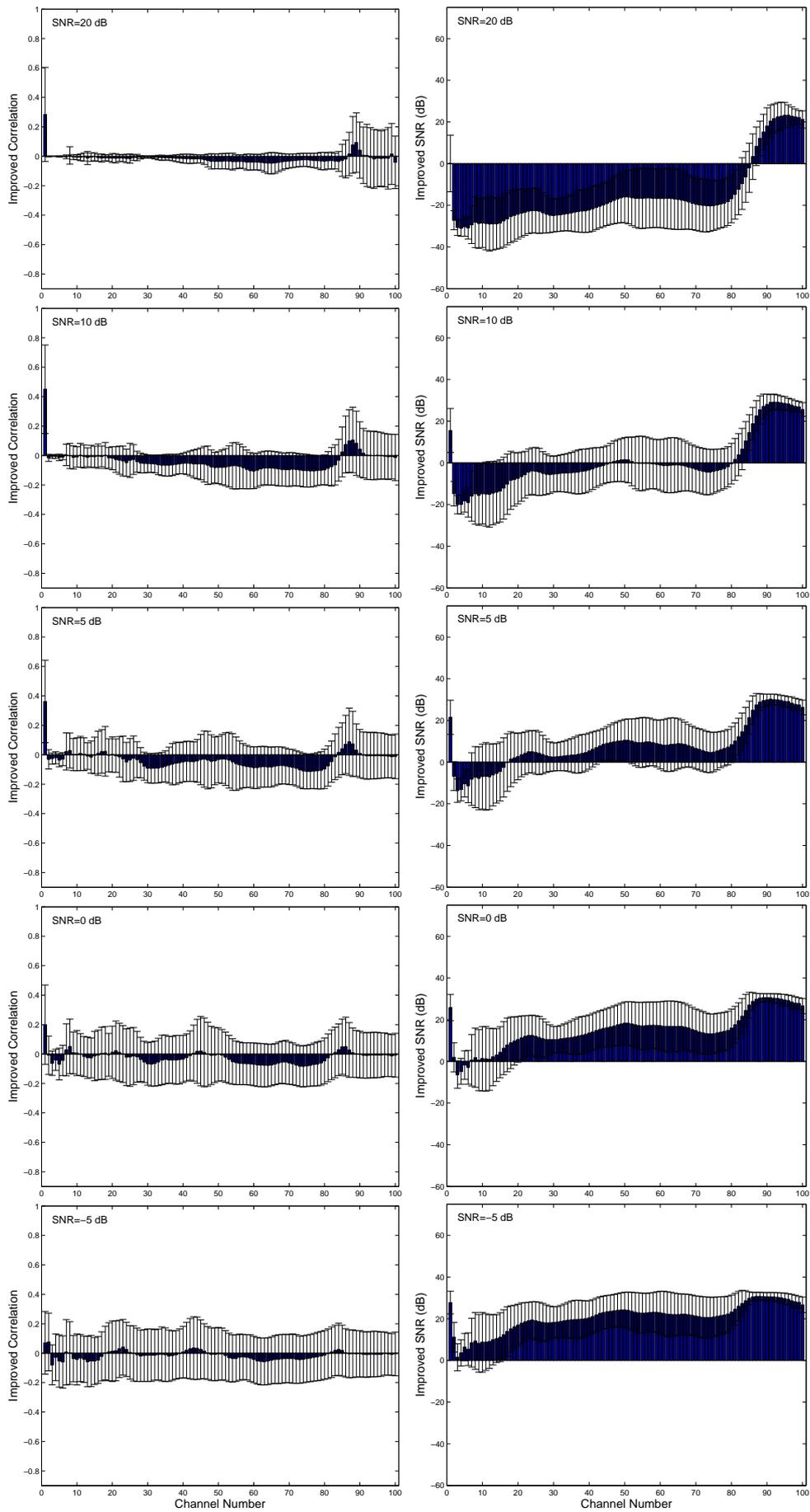


図 6.11: MMSE-STSA: ピンク雑音での相関値とパワー比の改善度

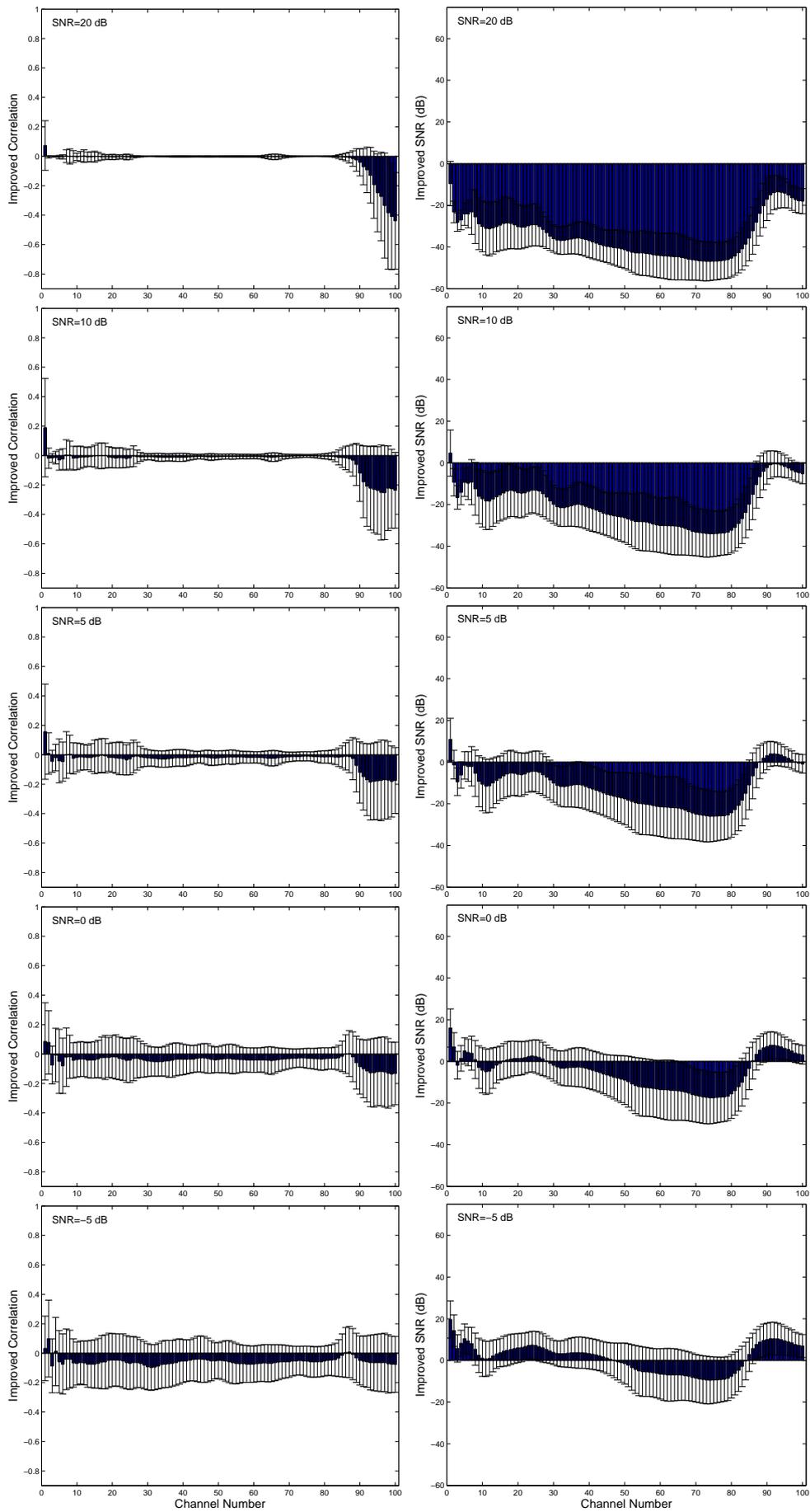


図 6.12: MMSE-STSA:バブル雑音での相関値とパワー比の改善度

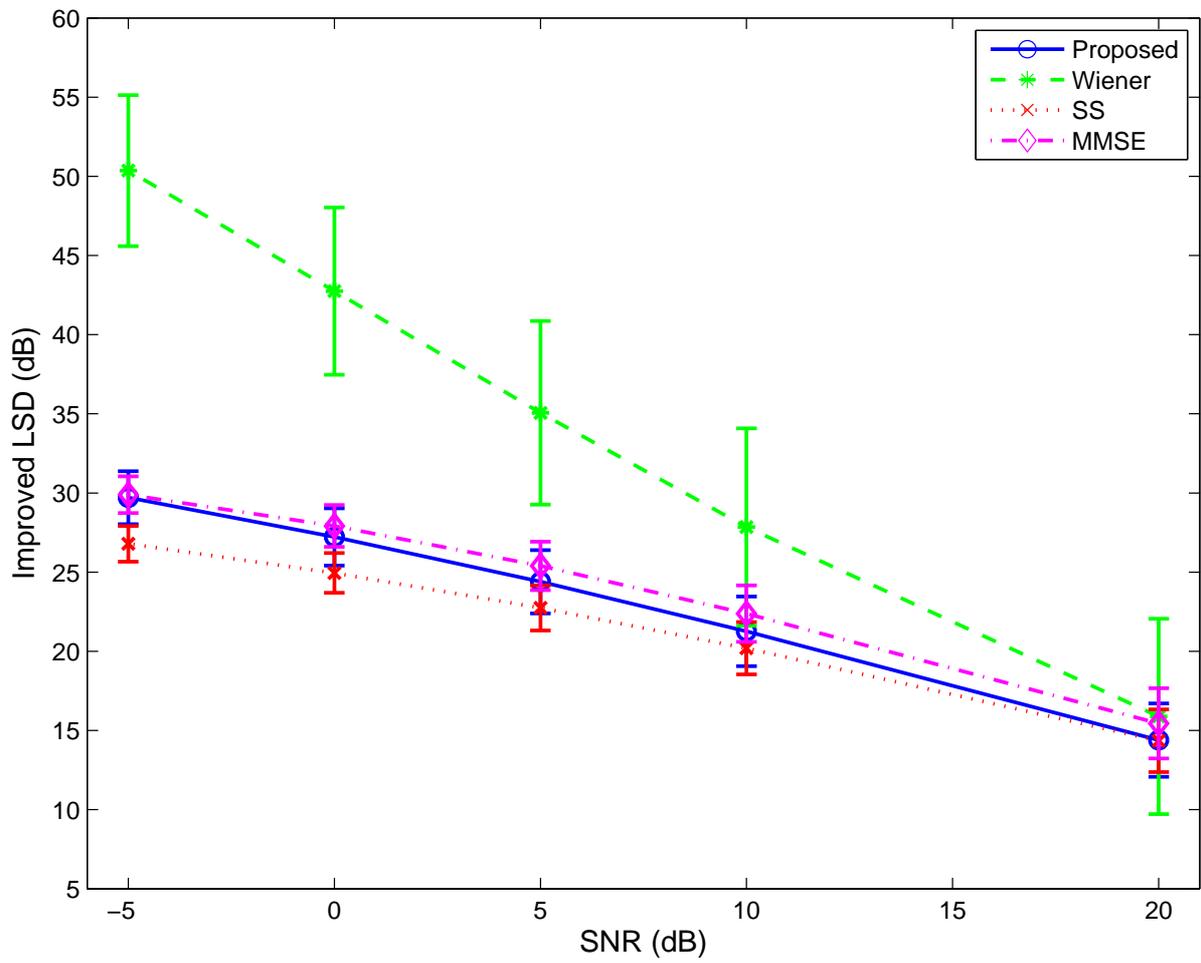


図 6.13: 白色雑音での LSD の改善度

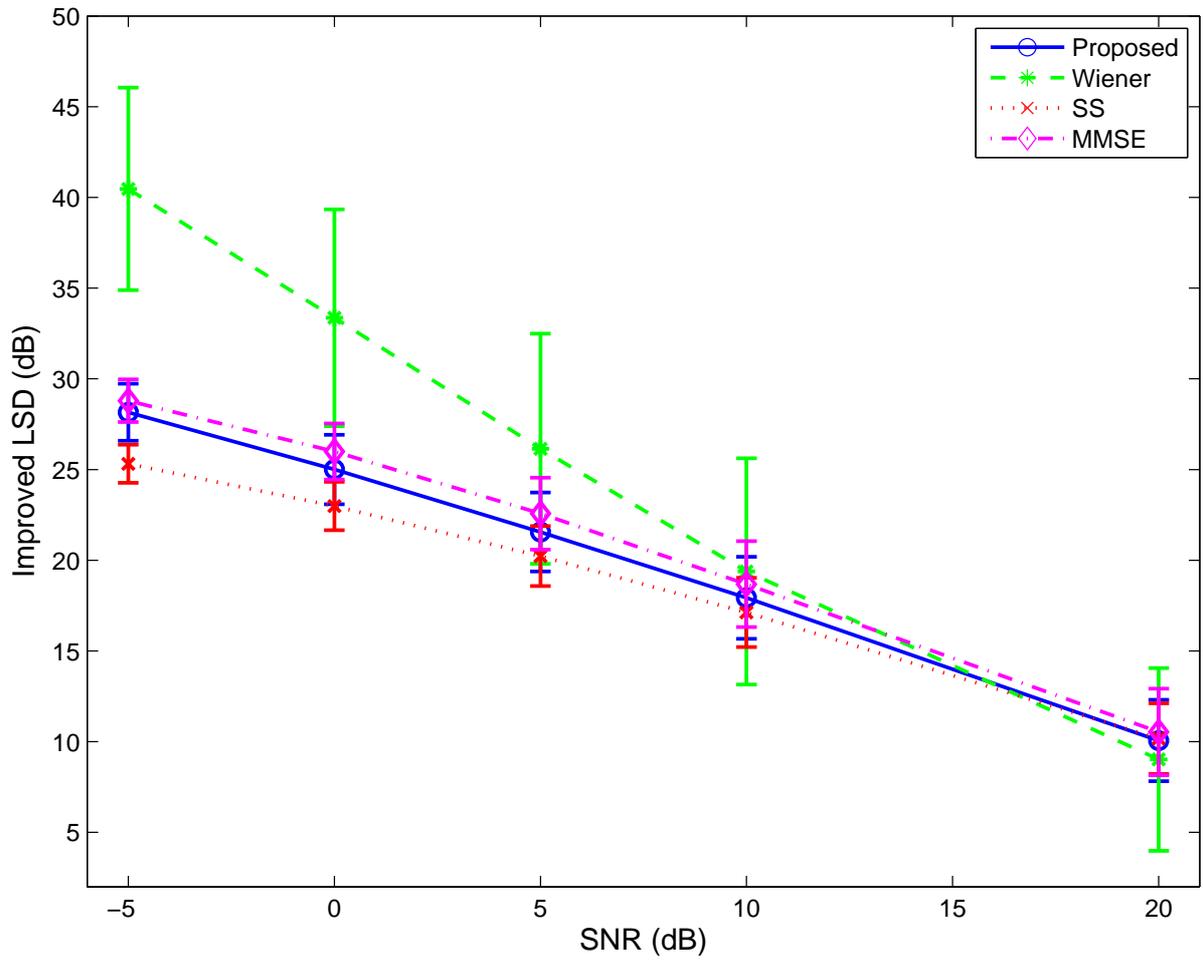


図 6.14: ピンク雑音での LSD の改善度

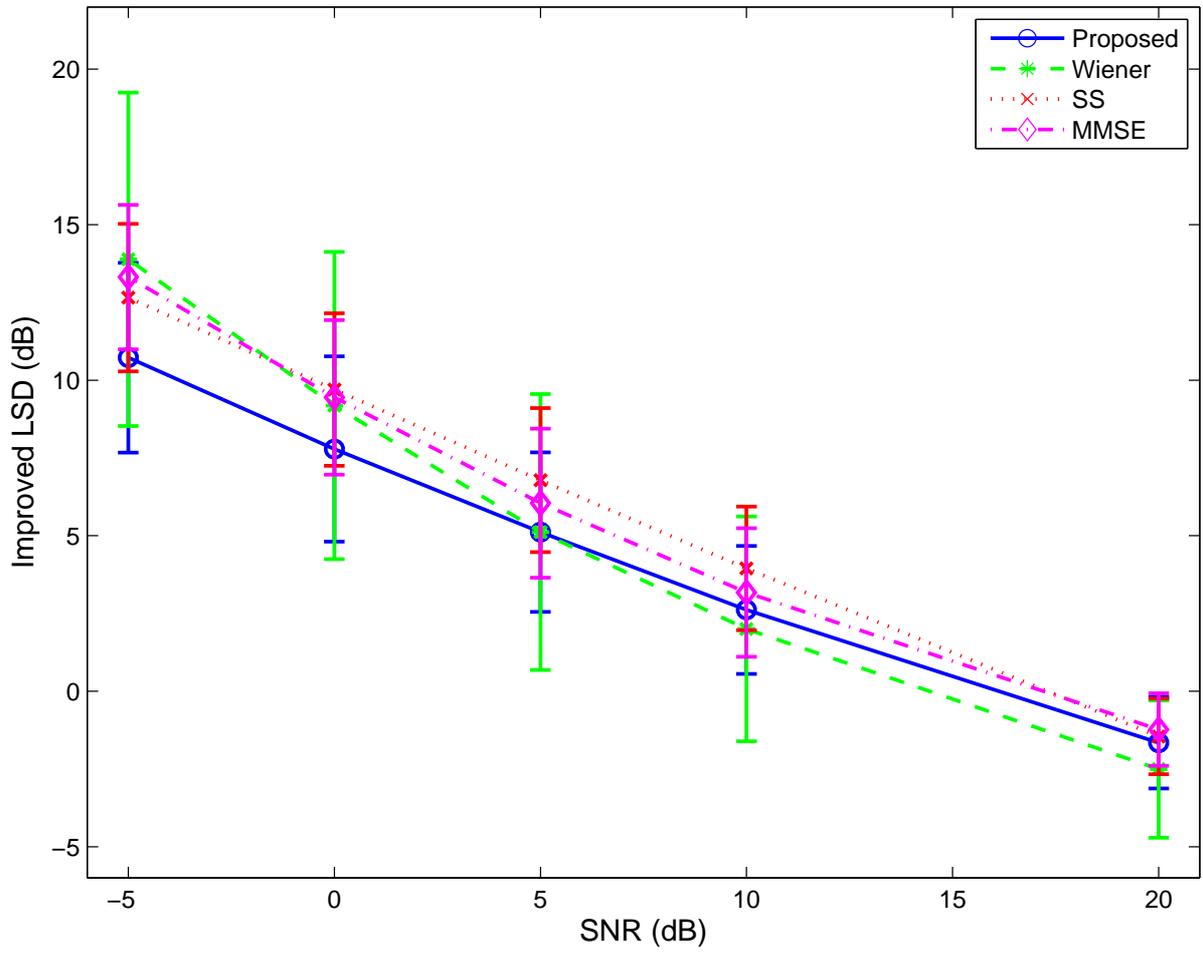


図 6.15: バブル雑音での LSD の改善度

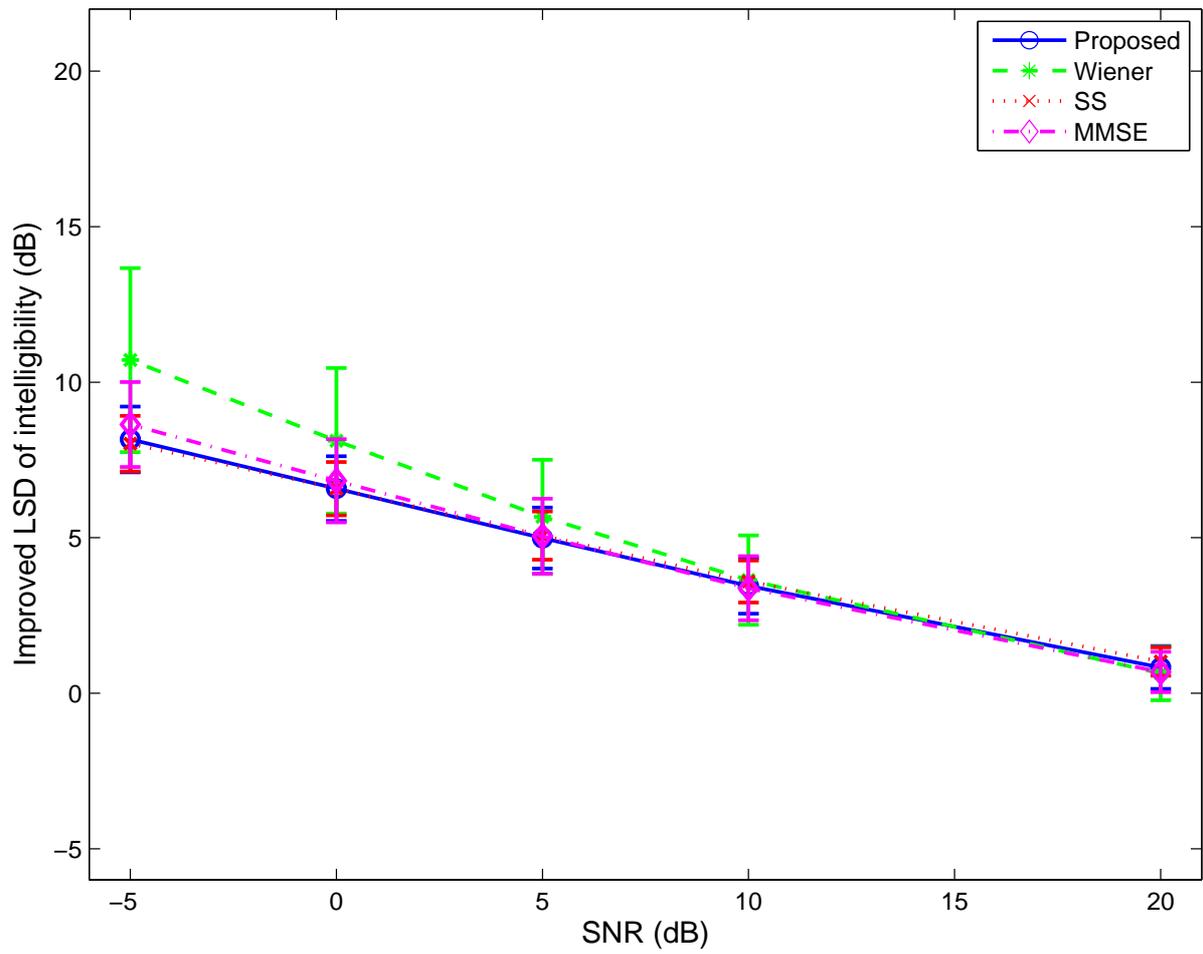


図 6.16: 白色雑音での重み付き LSD の改善度

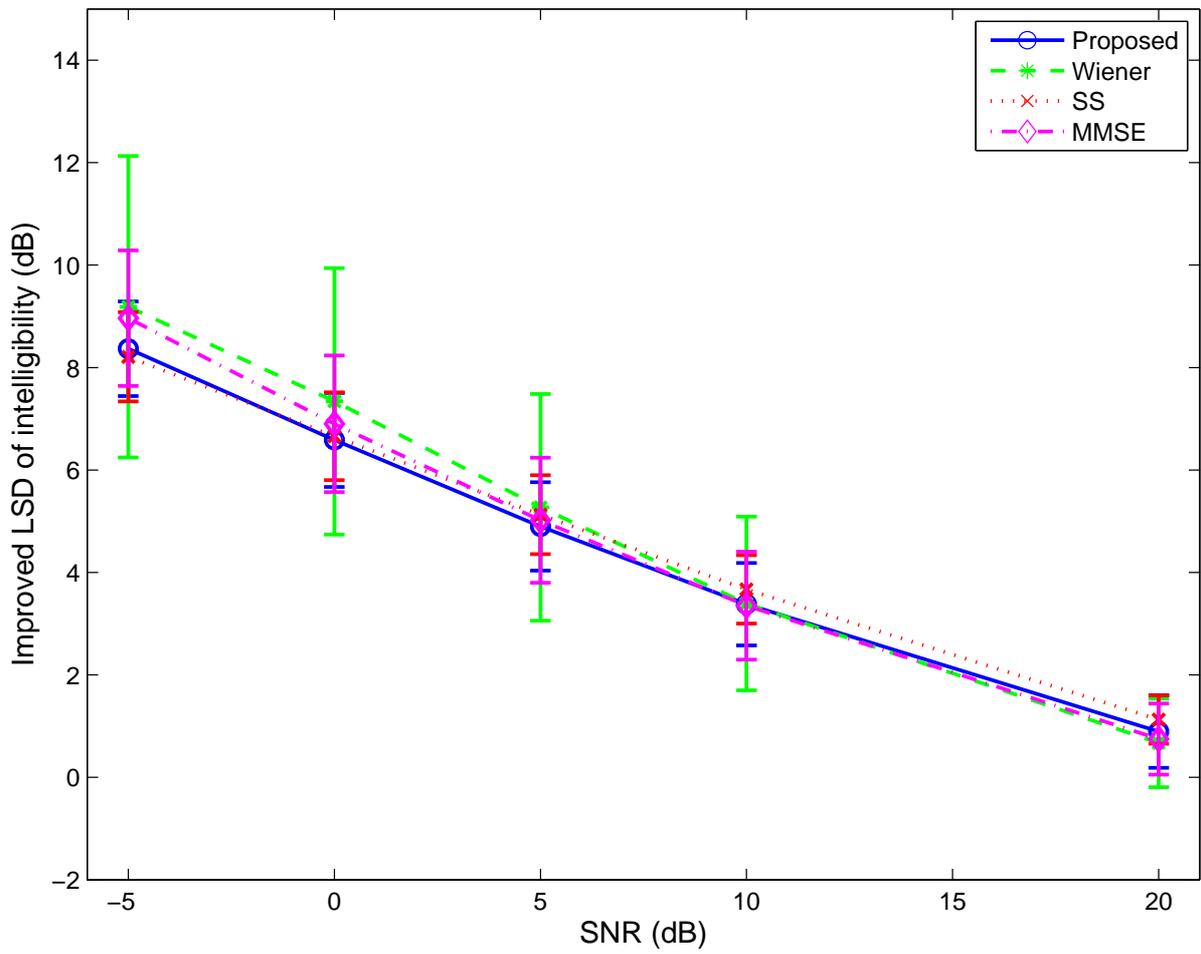


図 6.17: ピンク雑音での重み付き LSD の改善度

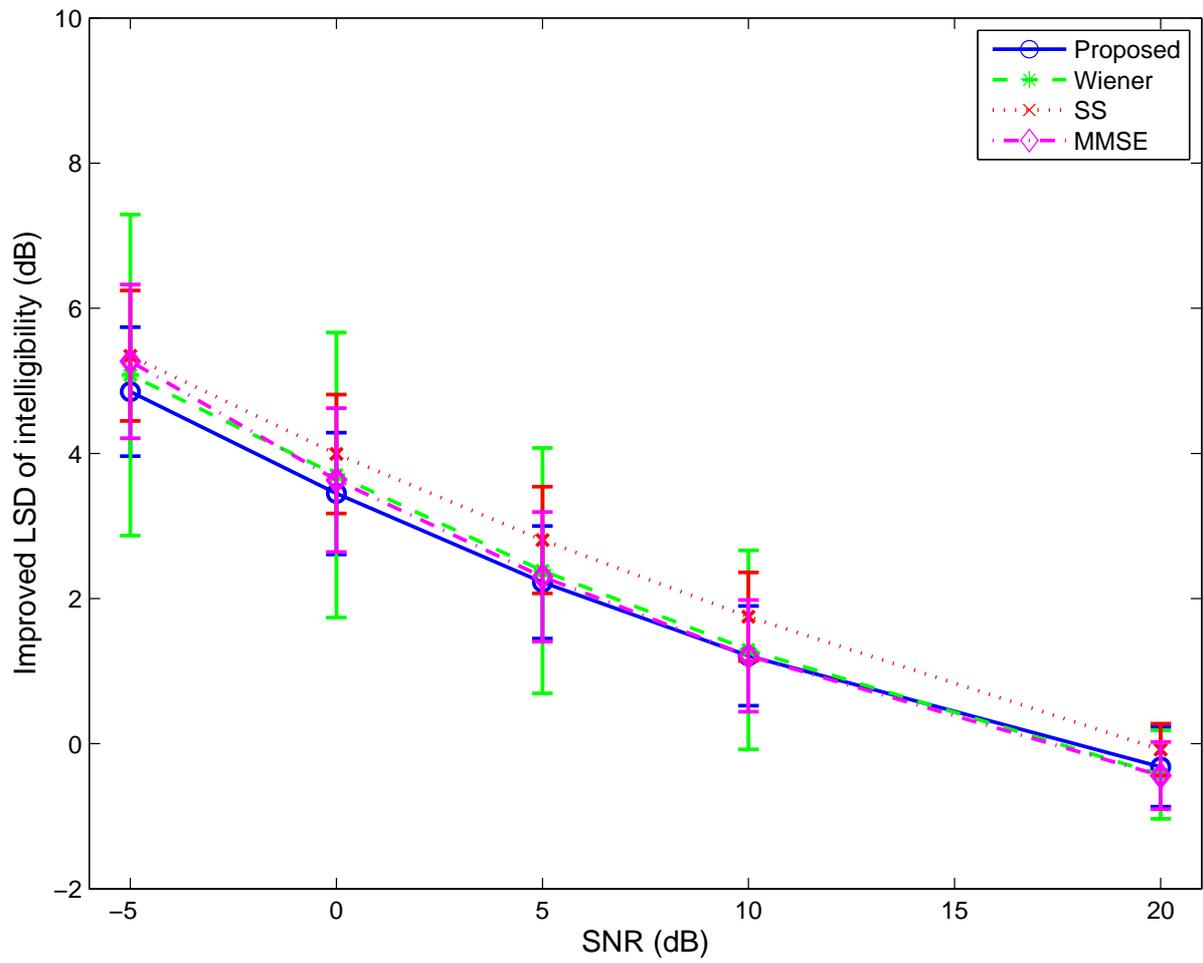


図 6.18: バブル雑音での重み付き LSD の改善度

第7章 結論

7.1 本研究で明らかにしたこと

本研究では，MTF の概念と Unoki らが提案したパワーエンベロープ逆フィルタ法の有用性を確認し，実現にまでは至っていないが，MTF に基づく雑音・残響抑圧法の基本となる処理体系を明らかにした．

雑音・残響抑圧法を実現するのに必要である，MTF に基づいた雑音抑圧法を提案する際の問題点を明らかにし，これらを解決したうえで，観測した雑音音声の情報のみから音声の回復処理を行なう MTF に基づいた雑音抑圧法を新たに提案した．

白色雑音，ピンク雑音，バブル雑音がそれぞれ付加された音声を用いて，提案手法を評価した結果，パワーエンベロープの相関値とパワー比の改善度より，パワーエンベロープを復元する点で雑音に対して効果があること，LSD の改善度より，音声信号を復元する点で効果があることを明らかにした．また，音声明瞭度と関係の取れた重み付き LSD の改善度より，音声明瞭度を復元していることを明らかにした．これにより，提案手法は雑音音声に対する雑音抑圧手法として有用であることが明らかになった．また，従来の雑音抑圧法との比較により，パワーエンベロープを復元する点で従来法より優れており，SS 法と MMSE-STSA 法に対して LSD，重み付き LSD の改善度の点で同等の効果が得られることを明らかにした．

7.2 本研究における今後の課題と展望

パワーエンベロープの回復精度向上

評価シミュレーションの結果から，提案法は，音声明瞭度に関係している MTF を回復しているにもかかわらず，LSD や音声明瞭度と関係が取られている重み付き LSD において，ウィナーフィルタリングに劣る傾向がみられた．これは，提案法を施すことで，客観的に音質や音声明瞭度の改善が行なわれたが，その効果が不十分であるからと考えられる．提案法の効果が不十分になってしまった原因は，雑音パワーエンベロープが時間領域で定常なのではなく，時間軸に対して変動しているためである．これにより，雑音パワーエンベロープの平均値より高い値をもった雑音成分が，提案法では抑えきれずに残ってしまっている．但し，SNR が 20，10 dB と高めの帯域では，雑音の動きを無視できる．この問題に対して，カルマンフィルタを用いることで，雑音パワーエンベロープの動きに対

応できる可能性がある。但し，パワーエンベロープに対して，MTFに基づいてカルマンフィルタ処理を行なう必要がある。

キャリア成分の回復

提案法は，パワーエンベロープの回復を行なっているが，キャリア成分の回復は行っていない。より精度良く音声を回復するためには，キャリア成分の回復を行なう必要がある。回復方法として，Unokiら [33] が提案している基本周波数を用いたキャリア再生成法を雑音環境に適応させることで，キャリア成分の回復ができる可能性がある。

雑音・残響抑圧法の提案

Unokiらの手法 [31]-[33] と提案法を組み合わせる事で，雑音・残響抑圧法の提案できる可能性がある。ただ単に提案法の後に Unoki らの手法を施した場合は，縦列的な処理になってしまい，雑音・残響の同時抑圧法の提案ができない。1つの見解として，式 (3.10) の MTF を逆フィルタリングすることで同時抑圧法を提案できると考える。ただし， $1/m(f_m)$ を掛ける際には，提案法と同様に，振幅に影響を及ぼさず，変調度のみに対して行なわれなければならないと考えられる。

音声認識システムや補聴システムへの応用

MTFに基づいている Unoki らの残響抑圧法 [33] は，音声明瞭度を最大で約 30 % ほど回復できることが分かっている。また，Lu ら [28] によって Unoki らの手法が残響環境下での音声認識の前処理に有効であることが分かっている。このことから，Unoki らの残響抑圧法は，音声認識システムや補聴システムに応用すれば，システムの精度向上へ寄与できると考えられる。残響環境か雑音環境の違いはあるが，Unoki らと同様に MTF に基づいている提案法も音声認識システムや補聴システムの精度向上に寄与できると考えられる。

謝辞

本研究を進めるにあたり，多大な助言と懇切丁寧かつ，熱心な御指導をして頂きました鶴木祐史准教授に心から感謝致します．本研究を進めるにあたり，多大な助言と熱心な指導をして頂きました赤木正人教授に心から感謝致します．本研究に関して多大な助言をして頂いた李軍峰助教，博士後期課程の羽二生篤氏，木谷俊介氏に心より感謝致します．有意義な討論，助言を賜った赤木・鶴木研究室の皆様方に心から感謝いたします．

参考文献

- [1] S. T. Neely, J. B. Allen, “Invertibility of a room impulse response,” *J. Acoust. Soc. Am.*, Vol. 66, No. 1, pp. 166–169, July 1979.
- [2] M. Miyoshi, Y. Kaneda, “Inverse filtering of room acoustics,” *IEEE Trans. ASSP.*, Vol. 36, No. 2, pp. 145–152, Feb. 1988
- [3] H. Wang, F. Itakura, “Realisation of acoustic inverse filtering through multi-microphone sub-band processing,” *IEICE Trans.*, Vol. E75-A No. 11, pp. 1474–1483.
- [4] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” *IEEE Trans. ASSP.*, Vol. 27, No. 2, pp. 113–120, 1979
- [5] 片山 徹, 応用カルマンフィルタ, 朝倉書店, 2000, pp. 66–82.
- [6] P. Scalart, J. V. Filho, “Speech enhancement based on a priori signal to noise estimation,” *ICASSP’96*, Vol. 2, pp. 629–632, 1996.
- [7] 石田 隆晃, 田口 亮, “データ依存型ウィナーフィルタの帯域分割処理による音声信号の雑音抑圧,” 電子情報通信学会技術研究報告, Vol. 105, No. 426, pp.19–24, 2005.
- [8] 西山 清, 中野 道雄, パソコンで解くカルマンフィルタ, 丸善, 1993 .
- [9] K. K. Paliwal, A. Basu, “A speech enhancement method based on Kalman filtering,” *ICASSP’87*, Vol. 1, pp. 177–180, 1987.
- [10] Y. Ephraim, and D. Malah, “Speech Enhancement using a minimum mean-square error short-time spectral amplitude estimator,” *IEEE Trans. ASSP.*, Vol. ASSP-32, No. 6, pp.1109-1121, Dec. 1984.
- [11] R. J. McAulay, M. L. Malpass, “Speech enhancement using a soft-decision noise suppression filter,” *IEEE Trans. ASSP.*, Vol. 28, No. 2, pp. 137–145, Apr. 1980.
- [12] 北岡 教英, 赤堀 一郎, 中川 聖一, “スペクトルサブトラクションと時間方向スムージングを用いた雑音環境下音声認識,” 電子情報通信学会論文誌, Vol. J83-D-2, No. 2, pp. 500–508, Feb. 2000.

- [13] 藤本 雅清, 有木 康雄, “カルマンフィルタに基づく音声信号推定法を用いた雑音環境下での音声認識,” 電子情報通信学会論文誌, Vol. J85-D-2 No. 1, pp. 1–11, Jan 2002.
- [14] 加藤 正徳, 杉山 昭彦, 芹沢 昌宏, “重み付き雑音推定と MMSE STSA 法に基づく高音質雑音抑圧,” 電子情報通信学会論文誌, Vol. J87-A, No. 7, pp. 851–860, Jul 2004.
- [15] 上村 益永, 高橋 祐, 猿渡 洋, 鹿野 清宏, “高次統計量を利用したミュージカルノイズ計量尺度に基づくスペクトル減算法の自動最適化,” 日本音響学会講演論文集, pp. 691–694, Sep. 2008.
- [16] 荒井 隆行, 木下 慶介, 程島 奈緒, 楠本 亜希子, 喜田村 朋子, “音声の定常部抑圧の残響に対する効果,” 日本音響学会講演論文集, pp. 449–450, Oct. 2001.
- [17] 吉田 航, 程島 奈緒, 荒井 隆行, “雑音・残響環境下における音声明瞭度改善のための定常部抑圧処理の評価,” 日本音響学会講演論文集, pp. 493–496, Sep. 2008.
- [18] K. Kinoshita, M. Delcroix, T. Nakatani, and M. Miyoshi, “Multi-step linear prediction based speech enhancement in noisy reverberant environment,” Proc. *Interspeech-2007.*, pp. 854–857, Aug. 2007.
- [19] 吉岡 拓也, 中谷 智広, 三好 正人, “雑音と残響の同時抑圧による音声強調,” 日本音響学会講演論文集, pp. 731–732, Mar. 2008.
- [20] 吉岡 拓也, 中谷 智広, 三好 正人, “雑音・残響抑圧を目的とした線形フィルタに非線形フィルタを後置させた系の最適化法,” 日本音響学会講演論文集, pp. 845–846, Sep. 2008.
- [21] 戸井田 義徳, “空間内における音声情報伝達,” 日本音響学会誌, 51 巻 4 号, pp. 312–316, 1995.
- [22] T. Houtgast and H. J. M. Steeneken, “The Modulation Transfer Function in Room Acoustics as a Predictor of Speech Intelligibility,” *Acoustica.*, Vol. 28, pp. 66–73, 1978.
- [23] T. Houtgast, H. J. M. Steeneken and R. Plomp, “Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. I. General Room Acoustics,” *Acoustica.*, Vol. 46, pp. 60–72, 1980.
- [24] T. Houtgast, H. J. M. Steeneken, “A review of the MTF concept in room acoustic and its use for estimating speech intelligibility in auditoria,” *J. Acoust. Soc. Am.*, Vol. 77, No. 3, March 1985.
- [25] 小椋 靖夫, 浜田 晴夫, 三浦 種敏, “音場における音声伝送品質のための MTF と STI について,” 日本音響学会誌, 40 巻 3 号, pp. 181–191, Mar. 1984.

- [26] 中島立視, “音声の明瞭度指標 (STI) の測定,” 日本音響学会誌, 49 巻 2 号, pp. 103–110, 1993.
- [27] 平松 壮太, 鷓木 祐史, “変調伝達関数に基づいた残響時間のブラインド推定法の検討,” 日本音響学会聴覚研究会資料, Vol. 37, No. 11, pp. 855–860, Jan. 2008.
- [28] L. Xugang, M. Unoki, and M. Akagi, “Comparative evaluation of modulation-transfer-function-based blind restoration of sub-band power envelopes of speech as a front-end processor for automatic speech recognition systems,” *Acoust. Sci. & Tech.*, Vol. 29, No. 6, pp. 351–361
- [29] 広林 茂樹, 野村 博昭, 小池 恒彦, 東山 三樹夫, “パワーエンベロープ伝達関数の逆フィルタ処理による残響音声の回復,” 電子情報通信学会論文誌, Vol. J81-A, No. 10, pp. 1323–1330, Oct. 1998.
- [30] M. Unoki, K. Sakata, and M. Akagi, “A speech dereverberation method based on the MTF concept,” *Eurospeech2003* in Geneva, pp. 1417–1420, 2003.
- [31] M. Unoki, M. Furukawa, K. Sakata, and M. Akagi, “An improved method based on the MTF concept for restoring the power envelope from a reverberant signal,” *Acoust. Sci. & Tech.* Vol. 25, No. 4, pp. 232–242, 2004.
- [32] M. Unoki, K. Sakata, M. Furukawa, and M. Akagi, “A speech dereverberation method based on the MTF concept in power envelope restoration,” *Acoust. Sci. & Tech.*, Vol. 25, No. 4, pp. 243–254. 2004.
- [33] M. Unoki, M. Toi, and M. Akagi, “Development of the MTF-based speech dereverberation method using adaptive time-frequency division,” *Proc. Forum Acusticum2005* in Budapest, pp. 51–56, 2005.
- [34] M. R. Schroeder, “Modulation transfer function: definition and measurement,” *Acoustica.*, Vol. 49, pp.179–182, 1981.
- [35] T. Arai, M. Pavel, H. Hermansky, and C. Avendano, “Syllable intelligibility for temporally filtered LPC cepstral trajectories,” *J.Acoust. Soc. Am.*, Vol. 105, No. 5, pp. 2783–2791, May 1999.
- [36] N. Kanedera, T. Arai, H. Hermansky, and M. Pavel, “On the importance of various modulation frequencies for speech recognition,” *Eurospeech97* in Rhodes, pp. 1079–1082, 1997.
- [37] 金寺 登, 荒井 隆行, 船田 哲男, “変調スペクトルの重要な成分のみを選択的に用いた雑音に強い音声認識,” 電子情報通信学会論文誌, Vol. 84, No. 7, pp. 1261–1269, Jul. 2001.

- [38] T. Takeda, Y. Sagisaka, K. Katagiri, M. Abe, and H. Kuwabara, *Speech Database User's Manual*, ATR Technical Report, TR-I-0028, 1988.
- [39] ANSI S3.5-1997, "American National Standard Methods for Calculation of the Speech Intelligibility Index," 1997.

発表

Y. Yamasaki, M. Unoki, “A study on the noise suppression method based on the MTF concept,” *2009 RISP International Workshop on Nonlinear Circuits and Signal Processing*, Mar 2009.