

Title	残響時間の自動推定による残響波形パワーエンベロープの回復に関する研究
Author(s)	古川, 正和
Citation	
Issue Date	2002-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1553
Rights	
Description	Supervisor: 赤木 正人, 情報科学研究科, 修士

修 士 論 文

残響時間の自動推定による残響波形パワーエンベ
ロープの回復に関する研究

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

古川 正和

2002 年 3 月

修 士 論 文

残響時間の自動推定による残響波形パワーエンベ
ロープの回復に関する研究

指導教官 赤木正人 教授

審査委員主査 赤木正人 教授
審査委員 党建武 助教授
審査委員 下平博 助教授

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

010103 古川 正和

提出年月: 2002 年 2 月

概 要

実環境で観測される信号には、音源信号の他に雑音や残響が混在している。そのため実環境下で音声信号処理を行う場合、その精度が著しく低下する。特に残響は、音声信号処理の性能を低下させる主な原因となる。現在、残響信号から音源信号を回復する逆フィルタ処理が数多く検討されており、その有効性が示されている。しかし、提案された手法は室内伝達特性のパラメータが既知であるという制約条件があるため、実環境下での適応は困難である。

広林らは、MTF(Modulation Transfer Function)の理論に基づき残響波形のパワーエンベロープ回復を目的としたパワーエンベロープ逆フィルタ処理を提案している。これにより、インパルス応答の測定をすることなく残響の影響を軽減できたが、残響時間が既知であるという制約条件が存在する。

本研究では、残響波形の情報のみから残響時間を自動推定してパワーエンベロープ逆フィルタ処理により残響波形のパワーエンベロープを回復することを目的とする。パワーエンベロープ逆フィルタ処理は、パワーエンベロープを抽出する精度に依存しているため、パワーエンベロープ抽出法を検討した。その結果、ヒルベルト変換により抽出する方法が最適であるとわかった。残響時間の推定は、回復したパワーエンベロープが負の値を持たないようにすることで推定できることがわかった。音声のパワーエンベロープを回復する場合、帯域ごとにパワーエンベロープが異なることから帯域分割をする必要性について述べた。また、残響音声を帯域分割して各帯域にパワーエンベロープ逆フィルタ処理を施すことにより、音声のパワーエンベロープが回復されることを示した。

目次

第1章	序論	1
1.1	背景	1
1.2	目的	1
1.3	本論文の構成	2
第2章	パワーエンベロープ逆フィルタ法	3
2.1	原理	3
2.1.1	変調伝達関数 (MTF)	3
2.1.2	パワーエンベロープ逆フィルタ	4
2.2	検証シミュレーション	6
2.2.1	検証結果	6
2.3	問題点	7
2.3.1	パワーエンベロープ抽出法	7
2.3.2	残響時間の決定	8
2.3.3	音声への適用	8
第3章	パワーエンベロープ抽出法	13
3.1	パワーエンベロープ抽出法の提案	13
3.1.1	抽出法の原理	13
3.1.2	低域通過フィルタ	16
3.2	抽出法の比較実験	16
3.2.1	比較実験の条件	16
3.2.2	検証結果	17
3.2.3	考察	18
第4章	残響時間の推定	22
4.1	変調に関する仮定	22
4.2	インパルス応答の定数に関する仮定	22
4.3	パワーエンベロープ逆フィルタの特徴	22
4.4	残響時間の推定方法	23
4.4.1	残響時間推定法の提案	23

4.4.2	推定法の原理	23
4.5	残響時間推定の比較実験	26
4.5.1	検証シミュレーション	26
4.5.2	検証結果	26
4.5.3	考察	26
第5章	音声への適用	30
5.1	音声を用いる問題点	30
5.1.1	音声の搬送波	30
5.1.2	音声と共変調	30
5.2	調波複合音の搬送波	31
5.2.1	検証シミュレーション	32
5.2.2	検証結果	32
5.2.3	考察	35
5.3	非共変調信号の回復	37
5.3.1	検証シミュレーション	37
5.3.2	検証結果	37
5.3.3	考察	41
5.4	帯域分割による回復実験	41
5.4.1	分割する帯域幅の決定	41
5.4.2	検証シミュレーション	43
5.4.3	検証結果	43
5.4.4	考察	43
第6章	まとめ	48
6.1	本研究で明らかにされたことの要約	48
6.2	今後の課題	48

目 次

2.1	パワーエンベロープ逆フィルタ処理による信号回復.(a) 原信号, (b) 残響時間 $T_R = 0.5$ [s] の残響信号,(c) 観測信号のパワーエンベロープを既知とした場合の回復信号,(d) 回復信号	9
2.2	パワーエンベロープ逆フィルタ処理によるパワーエンベロープ回復.(a) 原信号,(b) 残響時間 $T_R = 0.5$ [s] の残響信号,(c) 観測信号のパワーエンベロープを既知とした場合の回復信号,(d) 回復信号	10
2.3	パワーエンベロープ逆フィルタ処理による音声の回復.(a) 原音声, (b) 残響時間 $T_R = 0.5$ [s] の残響音声,(c) 観測信号のパワーエンベロープを既知とした場合の回復音声,(d) 回復音声	11
2.4	パワーエンベロープ逆フィルタ処理によるパワーエンベロープ回復.(a) 原信号,(b) 残響時間 $T_R = 0.5$ [s] の残響信号,(c) 観測信号のパワーエンベロープを既知とした場合の回復信号,(d) 回復信号	12
3.1	ヒルベルト変換による瞬時振幅の計算	14
3.2	低域フィルタの振幅特性	17
3.3	パワーエンベロープの抽出精度.(a) 相関値の平均値と標準偏差, (b) SNR の平均値と標準偏差	19
3.4	残響を付加したときの相関値の関係.(a) ヒルベルト変換による方法,(b) パワーによる方法,(c) 集合平均による方法,(d) 集合平均+パワーによる方法	20
3.5	残響を付加したときの SNR の関係.(a) ヒルベルト変換による方法, (b) パワーによる方法,(c) 集合平均による方法,(d) 集合平均+パワーによる方法	21
4.1	パワーエンベロープ逆フィルタの特徴	23
4.2	過剰な回復により現れる負の面積 (付加した残響時間 $T_R = 0.5$ [s]). 実線: 正確な回復のパワーエンベロープ ($\hat{T}_R = 0.5$), 点線: 回復不足のパワーエンベロープ ($\hat{T}_R = 0.3$), 破線: 過剰回復のパワーエンベロープ ($\hat{T}_R = 1.0$)	24
4.3	回復時に用いる残響時間と回復パワーエンベロープにおける周波数成分の関係	25
4.4	負の面積による残響時間の推定	28
4.5	変調伝達関数による残響時間の推定	28
4.6	高周波数成分による残響時間の推定	28

4.7	推定残響時間と相関,SNR の関係.(a) $T_R = 0.1$ [s], (b) $T_R = 0.3$ [s], (c) $T_R = 0.5$ [s], (d) $T_R = 1.0$ [s], (e) $T_R = 2.0$ [s].(上) 相関値,(下)SNR.	29
5.1	一括処理における音声のスペクトログラム.(a) 原音声,(b) 残響時間 $T_R = 0.5$ [s] の残響を付加した音声,(c) 観測した音声のパワーエンベロープを既知としたときの回復音声,(d) 残響時間を既知とした回復音声	31
5.2	搬送波が調波複合音のときのパワーエンベロープの抽出精度.	33
5.3	残響を付加したときの相関値と SNR の関係.(a) 相関値,(b)SNR	34
5.4	搬送波が調波複合音の場合の負の面積による残響時間の推定.(a) 残響信号パワーエンベロープの理論値を用いて推定した残響時間,(b) 残響信号からヒルベルト変換を用いて抽出したパワーエンベロープを用いて推定した残響時間	35
5.5	推定残響時間と相関,SNR の関係.(a) $T_R = 0.1$ [s], (b) $T_R = 0.3$ [s], (c) $T_R = 0.5$ [s], (d) $T_R = 1.0$ [s], (e) $T_R = 2.0$ [s]	36
5.6	非共変調信号.(a) 原信号,(b) 残響時間 $T_R = 0.5$ [s] の残響信号,(c) 各帯域における観測信号のパワーエンベロープを既知とした回復信号,(d) 帯域分割による回復信号,(e) 一括処理における観測信号のパワーエンベロープを既知とした回復信号,(f) 一括処理における回復信号	38
5.7	非共変調信号のスペクトログラム.(a) 原信号,(b) 残響時間 $T_R = 0.5$ [s] の残響信号,(c) 各帯域における観測信号のパワーエンベロープを既知とした回復信号,(d) 帯域分割による回復信号,(e) 一括処理における観測信号のパワーエンベロープを既知とした回復信号,(f) 一括処理における回復信号	39
5.8	各チャンネルにおけるパワーエンベロープの回復度.□:分割, なし:一括. 左から (1) 各帯域での残響時間を一定として帯域分割により回復した場合,(2) 各帯域で残響時間を推定して帯域分割により回復した場合, (3) 残響時間を既知として一括処理で回復した場合,(4) 残響時間を推定して一括処理で回復した場合,(5) 一括処理で、観測信号のパワーエンベロープを既知とした場合	40
5.9	各帯域間でのパワーエンベロープの相関値 (0.95 以上).	42
5.10	帯域分割により回復した音声のスペクトログラム (a) 原音声,(b) 残響時間 0.5 [s] の残響を付加した残響音声,(c) 全ての帯域で残響時間が同じとしたときの回復音声,(d) 残響時間を推定したときの回復音声, (d) 観測信号のパワーエンベロープを既知としたときの回復音声.	44
5.11	各チャンネルにおけるパワーエンベロープ回復指標.(a) 残響時間を一定とした場合,(b) 残響時間を推定した場合	45
5.12	各チャンネルにおけるパワーエンベロープ回復指標 (回復されない場合).(a) 残響時間を一定とした場合,(b) 残響時間を推定した場合	46

5.13	異なる搬送波白色雑音における抽出したパワーエンベロープ(全帯域).(a)MTF の理論が成立しているとき,(b)MTF の理論が成立していないとき (実線:残 響音声のパワーエンベロープ, 点線:原音声のパワーエンベロープ, 破線:残 響音声パワーエンベロープの理論値)	47
5.14	異なる搬送波白色雑音における抽出したパワーエンベロープ(1.6~2 [kHz]).(a)MTF の理論が成立しているとき,(b)MTF の理論が成立していないとき (実線:残 響音声のパワーエンベロープ, 点線:原音声のパワーエンベロープ, 破線:残 響音声パワーエンベロープの理論値)	47
6.1	ブロックダイアグラム.	50

第1章 序論

1.1 背景

実環境で観測される信号は、音源信号の他に雑音や残響が混在している。そのため実環境下で音声信号処理を行う場合、雑音や残響の影響によりその精度が著しく低下する。特に残響は、時間的に遅れた反射波が音源信号の直接波に大きなひずみを与えるため、音声信号処理の性能を低下させる主な原因となる。

現在、残響信号から音源信号を回復する逆フィルタ処理が数多く検討されており、その有効性が示されている。三好, 金田ら [1] は、音源の数に対しマイクロホンの一つ以上多く配置することでインパルス応答が最小位相特性を有してなくても音源波形そのものを正確に復元できる音場逆フィルタ理論を提唱している。それぞれの音源からマイクロホンに至る伝達関数の零点が重複していなければ、伝達系が最小位相を有していなくても音源信号を正確に復元できる。また王, 板倉 [2] らは、それぞれの帯域ごとに音源信号と回復信号の誤差が最小になる最適なインパルス応答の逆フィルタを求め音源波形そのものを復元する帯域分割逆フィルタ理論を提唱している。音源と複数のマイクロホン間におけるそれぞれの伝達系に対して、帯域ごとに音源信号と回復信号の誤差を最小にする最適な逆フィルタを求め音源信号を復元している。これらの処理は、逆フィルタを推定する上であらかじめ残響系のインパルス応答を測定しておかなければならない。しかし、残響系が推定されても室内伝達系は室内環境の変動を伴う時変系であるため、変動するごとに伝達系を測定しなければ回復精度が低下する。

広林らは、残響音はスペクトルひずみだけでなく波形の包絡線に影響を与えることに着目している。MTF (Modulation Transfer Function) の理論 [7] に基づき音源信号と伝達系をモデル化することで、音源信号そのものを復元するのではなくパワーエンベロープ回復を目的としたパワーエンベロープ逆フィルタ処理を提案している [3]。これにより、インパルス応答の測定をすることなく残響信号のパワーエンベロープを回復することができる。しかし、パワーエンベロープ逆フィルタ処理では残響時間が既知であるという制約条件が存在するため、実環境下において残響の影響を軽減することは困難である。

1.2 目的

本研究では、残響空間に存在するパラメータのうち残響時間パラメータを残響信号の情報のみから自動推定して、残響信号のパワーエンベロープを回復することを目的とする。

残響信号パワーエンベロープの回復には、パワーエンベロープ逆フィルタ処理を用いる。そこで、パワーエンベロープ逆フィルタが持つ問題点を明らかにする。明らかになった問題点を解決して残響音声のパワーエンベロープを回復することで、室内残響の影響を軽減する可能性を検討する。

本研究の特色は残響空間におけるパラメータのうち残響時間に着目して、残響時間パラメータを残響信号の情報のみから自動推定することである。残響時間パラメータの推定を逐次行い音源信号のパワーエンベロープを回復することで、室内伝達特性が時変系である空間や音源が移動している場合においても有効に残響を抑圧する手法が構築できる。

1.3 本論文の構成

本論文は全6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抑圧、パワーエンベロープ逆フィルタ処理と目的について述べる。

第2章では、パワーエンベロープ逆フィルタ処理における原理の説明をする。そして、パワーエンベロープ逆フィルタ処理の有効性と問題点について述べる。

第3章では、第2章で述べた問題点であるパワーエンベロープ抽出法について7つの手法を提案する。提案した手法を比較して、パワーエンベロープ抽出法として最適な手法を求める。

第4章では、本研究の特色でもある残響時間の推定について述べる。残響時間推定法を3つ提案する。残響時間推定法として最適な手法を求め、その有効性を検証する。

第5章では、パワーエンベロープ逆フィルタ処理を音声に適用するために、音声を模擬した信号を用いて提案法の検証をする。そして残響音声のパワーエンベロープを回復する。

第6章では、本研究で得られた結果を要約すると共に、今後の課題について述べる。

第2章 パワーエンベロープ逆フィルタ法

広林らは、振幅の特徴を反映する MTF (Modulation Transfer Function: 変調伝達関数) の理論に基づき音源信号と伝達系をモデル化することで、音源信号のパワーエンベロープ回復を目的としたパワーエンベロープ逆フィルタ処理を提案している [3]。次にパワーエンベロープ逆フィルタ処理の原理を述べる。

2.1 原理

2.1.1 変調伝達関数 (MTF)

Houtgast[7] らによって提案されている MTF による音声明瞭度の予測理論は、音声のパワーエンベロープひずみを評価するものである。

MTF と室内インパルス応答との関係は

$$m(\omega) = \int_0^\infty h^2(t) e^{j\omega t} dt / \int_0^\infty h^2(t) dt \quad (2.1)$$

と表される。つまり、 $m(\omega)$ は 2 乗室内インパルス応答の複素フーリエ変換を全エネルギーで除算したものである。 $h(t)$ を

$$h(t) = e^{-t/\tau} \quad (2.2)$$

とすると、式 (2.1) より $m(\omega)$ は以下のように表される。

$$m(\omega) = 1/(1 + j\omega\tau/2) \quad (2.3)$$

$$|m(\omega)| = \left[1 + \left(\omega \frac{\tau}{2} \right)^2 \right]^{-1/2} \quad (2.4)$$

室内インパルス応答は、確率的に次のようにモデル化できる。

$$h(t) = e^{-t/\tau} w(t) \quad (2.5)$$

ここで $w(t)$ はある確率過程から生じた実現値であり、白色雑音である。

式 (2.5), (2.1) より残響環境下における変調伝達関数を以下のように定義できる、

$$m(\omega) = \left[1 + \left(\omega \frac{T_R}{13.8} \right)^2 \right]^{-1/2} \quad (2.6)$$

ここで $\tau/2$ は $T_R/13.8$ に置き換えられている。また、 T_R はエネルギーが 60 [dB] 減衰する残響時間である。

これよりパワーエンベロープのひずみが室内の残響音による影響である場合、MTF は室内のインパルス応答から求めることができる [6]。

2.1.2 パワーエンベロープ逆フィルタ

音源信号を $x(t)$ 、室内インパルス応答を $h(t)$ とすると、残響信号 (観測信号) $y(t)$ は $x(t)$ と $h(t)$ の畳込みで表すことができる。

$$y(t) = x(t) * h(t) \quad (2.7)$$

広林らは MTF の理論に基づき、音源信号と室内インパルス応答を以下のようにモデル化している。

$$x(t) = e_x(t)n_1(t) \quad (2.8)$$

$$h(t) = e_h(t)n_2(t) \quad (2.9)$$

$$e_h(t) = ae^{\frac{-6.9t}{T_R}} \quad (2.10)$$

ここで、 $e_x(t)$ 、 $e_h(t)$ はそれぞれの包絡線であり、 a は定数、 T_R [s] は残響時間 (Reverberation time) である。 $n_1(t)$ 、 $n_2(t)$ は互いに無相関な白色雑音であり、

$$\langle \mathbf{n}_k(t)\mathbf{n}_k(t - \tau) \rangle = \delta(\tau) \quad (2.11)$$

が成立する。ここで、 $\mathbf{n}_1$ 、 $\mathbf{n}_2$ は白色ガウス過程より生起される白色雑音変数であり、 $n_1(t)$ 、 $n_2(t)$ はそれぞれ生起された実現値である。

今、ある確率過程から生成される残響信号 $y(t)$ の 2 乗集合平均を考える。残響信号が原信号と室内インパルス応答の畳込みで表されることから

$$\langle y(t)^2 \rangle = \left\langle \left\{ \int_{-\infty}^{\infty} x(\tau)h(t - \tau)d\tau \right\}^2 \right\rangle \quad (2.12)$$

$$= \left\langle \int_{-\infty}^{\infty} e_x(\tau_1)\mathbf{n}_1(\tau_1)e_h(t - \tau_1)\mathbf{n}_2(t - \tau_1) \int_{-\infty}^{\infty} e_x(\tau_2)\mathbf{n}_1(\tau_2)e_h(t - \tau_2)\mathbf{n}_2(t - \tau_2)d\tau_1d\tau_2 \right\rangle \quad (2.13)$$

のように求められる。但し、 $\langle \cdot \rangle$ は集合平均である。 $\langle y(t) \rangle$ は期待値の計算であるので

$$\begin{aligned} \langle y(t)^2 \rangle &= \int_{-\infty}^{\infty} e_x(\tau_1)e_h(t - \tau_1) \int_{-\infty}^{\infty} e_x(\tau_2)e_h(t - \tau_2) \\ &\quad \langle \mathbf{n}_1(\tau_1)\mathbf{n}_1(\tau_2) \rangle \langle \mathbf{n}_2(t - \tau_1)\mathbf{n}_2(t - \tau_2) \rangle d\tau_1d\tau_2 \end{aligned} \quad (2.14)$$

となる。式 (2.11) の関係から

$$\begin{aligned}\langle y(t)^2 \rangle &= \int_{-\infty}^{\infty} e_x(\tau_1) e_h(t - \tau_1) \int_{-\infty}^{\infty} e_x(\tau_2) e_h(t - \tau_2) \\ &\quad \delta(\tau_2 - \tau_1)^2 d\tau_1 d\tau_2\end{aligned}\quad (2.15)$$

$$\begin{aligned}&= \int_{-\infty}^{\infty} e_x(\tau)^2 e_h(t - \tau)^2 d\tau \\ &= e_x(t)^2 * e_h(t)^2\end{aligned}\quad (2.16)$$

が求められる。 $\langle y(t)^2 \rangle$ はパワーエンベロープの集合平均であることから、

$$\begin{aligned}e_y(t)^2 &= \langle y(t)^2 \rangle \\ &= e_x(t)^2 * e_h(t)^2\end{aligned}\quad (2.17)$$

と表すことができる¹。これより、室内インパルス応答パワーエンベロープの逆特性を用いることで音声のパワーエンベロープを回復することができる。そこで、式 (2.10) より

$$e_h(t)^2 = a^2 e^{\frac{-13.8t}{T_R}}\quad (2.18)$$

のように室内インパルス応答のパワーエンベロープを近似すれば、その逆特性を持つパワーエンベロープ逆フィルタは室内の残響時間 (T_R) から求めることができる。サンプリング周期を T_s [s] とすれば $t > 0$ の式 (2.18) に対するパワーエンベロープの伝達特性 $P_h(z)$ は

$$\begin{aligned}P_h(z) &= a^2 + a^2 \alpha z^{-1} + a^2 \alpha^2 z^{-2} + a^2 \alpha^3 z^{-3} + \dots \\ &= \frac{a^2}{1 - \alpha z^{-1}}\end{aligned}\quad (2.19)$$

但し, $\alpha = e^{\frac{13.8T_s}{T_R}}$

と表される。従って、音声のパワーエンベロープ特性 $P_x(z)$ は

$$\begin{aligned}P_x(z) &= \frac{P_y(z)}{P_h(z)} \\ &= \frac{1}{a^2} (1 - \alpha z^{-1}) P_y(z)\end{aligned}\quad (2.20)$$

$$\alpha = \exp\left(\frac{-13.8T_s}{T_R}\right)\quad (2.21)$$

のように残響音声のパワーエンベロープ特性 $P_y(z)$ とパワーエンベロープ伝達関数 $P_h(z)$ の逆特性の積で表される。 T_s はサンプリング周波数、 $P_x(z)$, $P_y(z)$ はそれぞれ $e_x(t)^2$, $e_y(t)^2$ の z 変換である。 $e_x(t)^2$ と $e_y(t)^2$ の正の平方根 (すなわち振幅包絡) を用いて、復元音声 $\hat{x}(t)$ は

$$\hat{x}(t) = \hat{e}_x(t) \frac{y(t)}{e_y(t)}\quad (2.22)$$

で表される。但し、右辺の $\hat{e}_x(t)$ は逆 z 変換で求められた回復エンベロープである。

¹以後、式 (2.17) により求めたパワーエンベロープを残響信号パワーエンベロープの理論値と呼ぶ。

2.2 検証シミュレーション

前節ではパワーエンベロープ逆フィルタ処理について説明をした。次に、パワーエンベロープ逆フィルタ処理の検証を行い、パワーエンベロープ回復精度の検証を行う。

2.2.1 検証結果

原信号の振幅エンベロープ $e_x(t)$ は

$$e_x(t) = \sqrt{1 + \sin 2\pi F t} \quad (2.23)$$

により作成する。ここで、 $F = 10$ [Hz] である。振幅エンベロープに搬送波である白色雑音を掛け合わせ原信号とする。残響信号は式 (2.7) により作成する。室内インパルス応答は式 (2.9) により求め、残響時間は 0.5 [s] とした。回復処理では残響時間を既知としている。

検証結果を図 2.1, 2.2 に示す。図 2.1(a) に原信号 $x(t)$ を示す。 $x(t)$ に残響時間 $T_R = 0.5$ [s] の室内インパルス応答を畳み込むことで図 2.1(b) の観測信号 $y(t)$ を得る。次に、この観測信号に対してパワーエンベロープ抽出処理を施すことによりパワーエンベロープを得る。図 2.2(a) および (b) にそれぞれ原信号 $x(t)$ と観測信号 $y(t)$ のパワーエンベロープを示す。なお、図 2.2(b) の破線は、式 (2.17) から求められた残響信号パワーエンベロープの理論値である。パワーエンベロープの抽出には、広林らが暫定的に用いているヒルベルト変換後の絶対値信号をカットオフ周波数 20 [Hz] で低域フィルタリングして求めている [4]。図 2.2 の (c) には残響信号パワーエンベロープの理論値から回復したパワーエンベロープを、(d) に抽出した残響信号パワーエンベロープから回復したパワーエンベロープをそれぞれ示す。また、図 2.1(c), (d) にそれぞれの回復信号を示す。

図 2.1, 2.2 の (c) から明らかなように、残響信号パワーエンベロープの理論値が抽出できれば、逆フィルタ処理によりほぼ完全にパワーエンベロープが回復できることがわかる。パワーエンベロープ抽出処理を施した残響信号のパワーエンベロープを用いた図 2.1, 2.2 の (d) では、回復パワーエンベロープに差異が生じており回復信号の概形も歪んでいる。これより、パワーエンベロープを抽出する精度が逆フィルタ処理によるパワーエンベロープの回復に影響を与えることがわかる。

次に、音声信号に対してパワーエンベロープ逆フィルタ処理をする。原音声のデータは ATR 音声データベース mau の“相変わらず”を用いた。残響時間は $T_R = 0.5$ [s] とした。図 2.3 にその結果を示す。また、図 2.4 にそれぞれのパワーエンベロープを示す。これらの結果から変調信号を用いた検証と同様に、残響音声パワーエンベロープの理論値が抽出できれば、逆フィルタ処理によりほぼ完全にパワーエンベロープを回復することができる。図 2.4 の (d) において、パワーエンベロープの回復効果がやや低いように見える。残響音声から抽出したパワーエンベロープと残響音声パワーエンベロープの理論値との差が大きいと、逆フィルタ処理により良好に回復ができないと考えられる。時間波形では、

残響による位相変化を全く考慮していないため完全な復元は不可能であり、回復音声の概観もひずみが見られる。

広林らはパワーエンベロープの回復を評価するために、パワーエンベロープ回復指標 (Improvement index of power envelope distortion: Ip) を式 (2.24) のように定義して定量的に評価している。

$$I_p = 10 \log_{10} \frac{\int_0^T \{e_x(t)^2 - e_y(t)^2\}^2 dt}{\int_0^T \{e_x(t)^2 - \hat{e}_x(t)^2\}^2 dt} \quad [\text{dB}] \quad (2.24)$$

ここで、 T は分析時間であり、 $T = 1$ [s] とした。また、 $\hat{e}_x(t)^2$ は回復したパワーエンベロープである。パワーエンベロープでの残響回復効果があればこの指標は高くなる。

変調信号を用いた場合の図 2.2(c),(d) において、式 (2.24) により測定したパワーエンベロープ回復指標は、それぞれ 234.29 [dB], 1.77 [dB] であった。また、音声を用いた場合の図 2.2(c),(d) ではそれぞれ 232.56 [dB], 0.25 [dB] であった。その結果、残響信号のパワーエンベロープを精度良く抽出できれば、パワーエンベロープ逆フィルタにより残響の回復が可能であることが示された。但し音声の場合、パワーエンベロープ抽出が変調信号よりも難しく、その影響が回復信号にも現れている。

2.3 問題点

パワーエンベロープ逆フィルタ処理を検証した結果、残響信号のパワーエンベロープを精度良く抽出できれば、残響信号のパワーエンベロープを回復する可能性があることがわかった。また、音声信号を変調信号としてみることであれば、パワーエンベロープ逆フィルタにより残響音声のパワーエンベロープが回復できると考えられる。しかし、検証シミュレーションを行う中で同時にいくつかの問題点が判明した。以下にそれらを示す。

1. パワーエンベロープの抽出方法
2. 残響時間の決定方法
3. 音声へ適用するときの仮定と問題点の考察

次にこれらの問題点について具体的に述べる。

2.3.1 パワーエンベロープ抽出法

図 2.2, 2.4 の (c) からわかるように、観測信号のパワーエンベロープが理論通りに抽出できれば、パワーエンベロープ逆フィルタ処理によりほぼ完全にパワーエンベロープを回復できる。これは、パワーエンベロープ逆フィルタ処理による回復精度が残響信号のパワーエンベロープを抽出する精度に依存していることを示している。そのため、残響信号からパワーエンベロープを抽出するとき精度良く抽出する必要がある。広林らは暫定的にヒル

ベルト変換を用いており、パワーエンベロープの抽出法についての考察が不十分である。本研究ではパワーエンベロープ抽出法について考察を行い、精度の高い手法を用いることにする。

2.3.2 残響時間の決定

広林らは、音源信号のパワーエンベロープを回復するパワーエンベロープ逆フィルタを提案している。これは伝達系の測定なしに残響を抑圧するものである。しかし、室内インパルス応答測定の必要はないが残響時間が既知でなければならない。残響時間は音源位置や室内環境などにより変化するため、実環境下においてパワーエンベロープ逆フィルタ処理を適用することは困難であると考えられる。本研究では、観測した残響信号の情報のみから回復することを目的としているため、残響時間の自動推定を行う手法を提案して、その検証を行う。

2.3.3 音声への適用

広林らは、パワーエンベロープ逆フィルタを音声に適用して残響抑圧の可能性を示している。しかし、原音声はMTFの理論に基づき振幅エンベロープと白色雑音でモデル化されているため、実際の音声信号との間には食い違いが生じる。この食い違いによる影響や問題点を検証する必要がある。本研究では、音声に適用する際の問題を明確にして検証を行う。また、音声信号の残響抑圧に関して定量的に評価を行う。

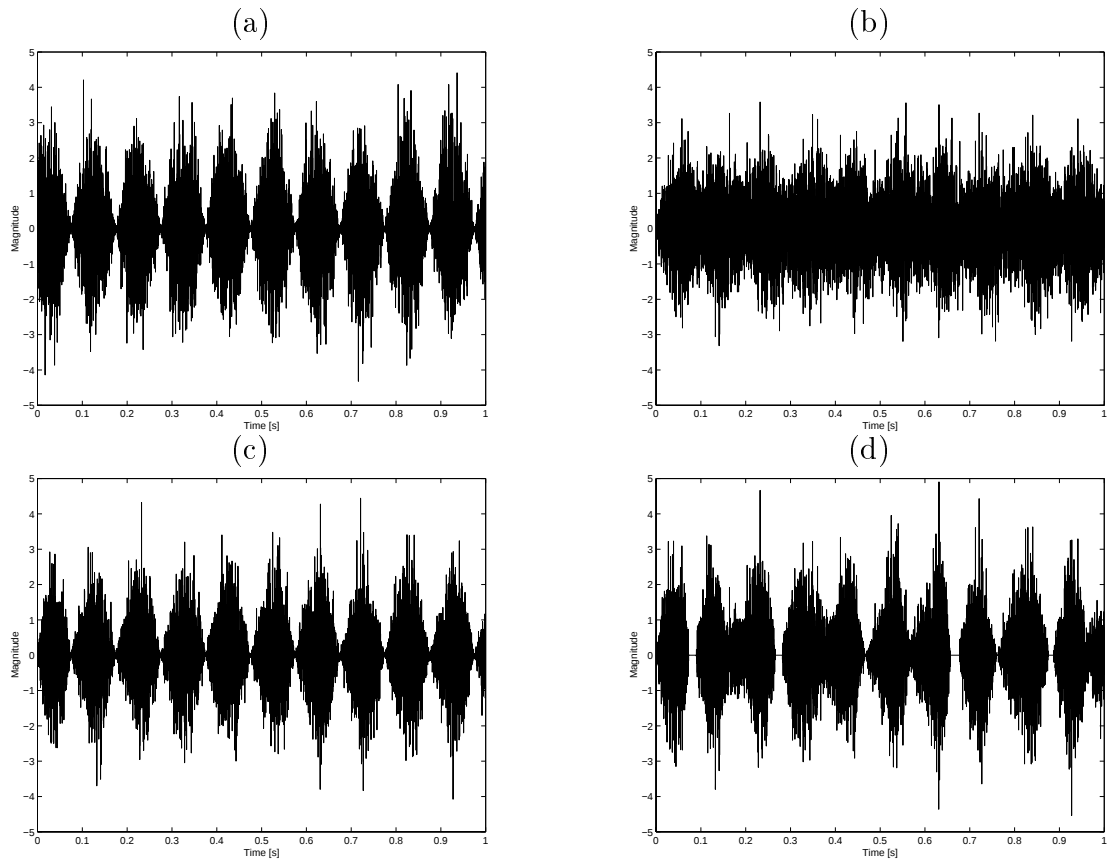


図 2.1: パワーエンベロープ逆フィルタ処理による信号回復.(a) 原信号, (b) 残響時間 $T_R = 0.5$ [s] の残響信号,(c) 観測信号のパワーエンベロープを既知とした場合の回復信号,(d) 回復信号

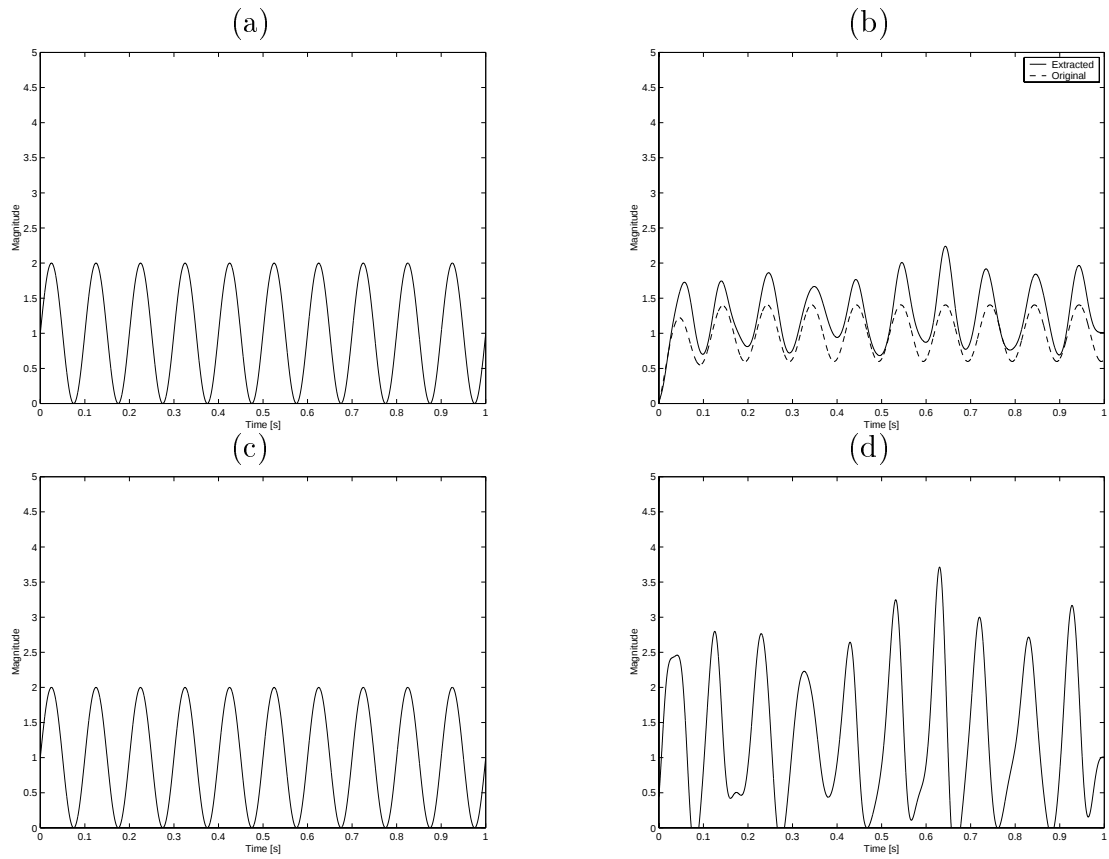


図 2.2: パワーエンベロープ逆フィルタ処理によるパワーエンベロープ回復.(a) 原信号,(b) 残響時間 $T_R = 0.5$ [s] の残響信号,(c) 観測信号のパワーエンベロープを既知とした場合の回復信号,(d) 回復信号

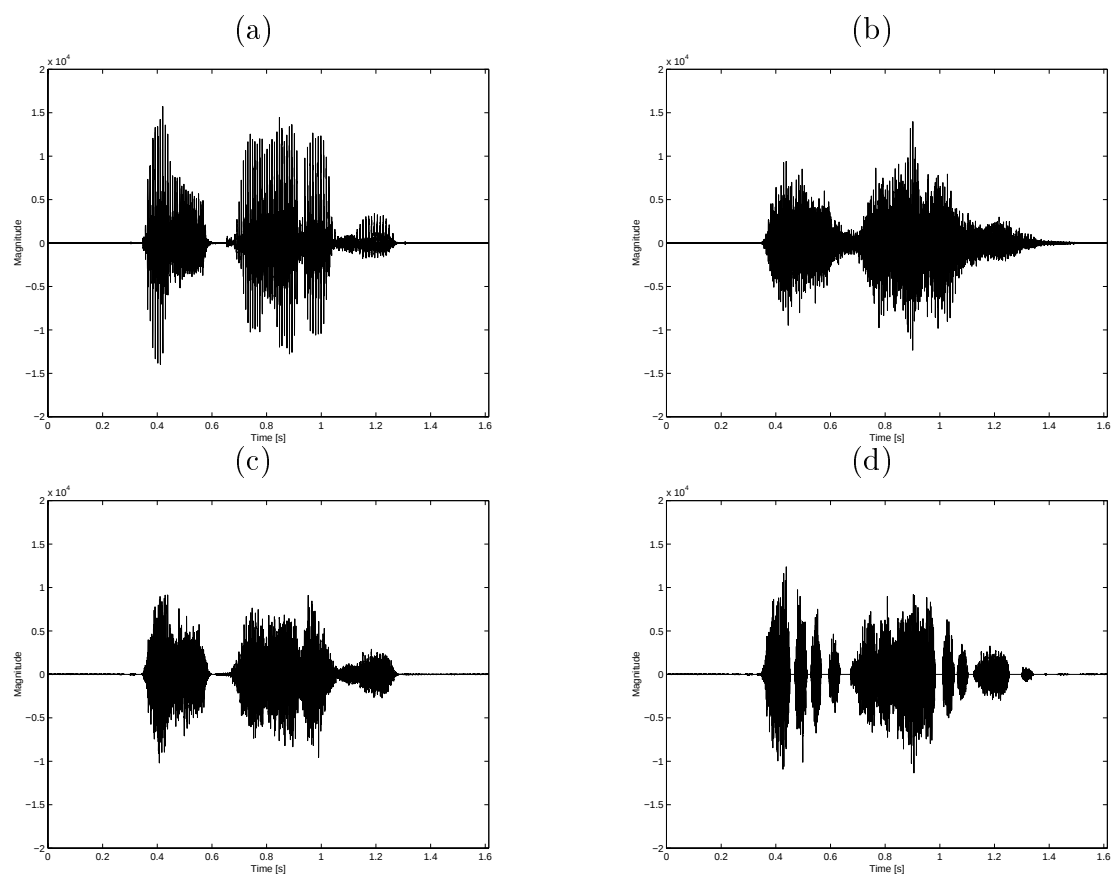


図 2.3: パワーエンベロープ逆フィルタ処理による音声の回復.(a) 原音声, (b) 残響時間 $T_R = 0.5$ [s] の残響音声,(c) 観測信号のパワーエンベロープを既知とした場合の回復音声,(d) 回復音声

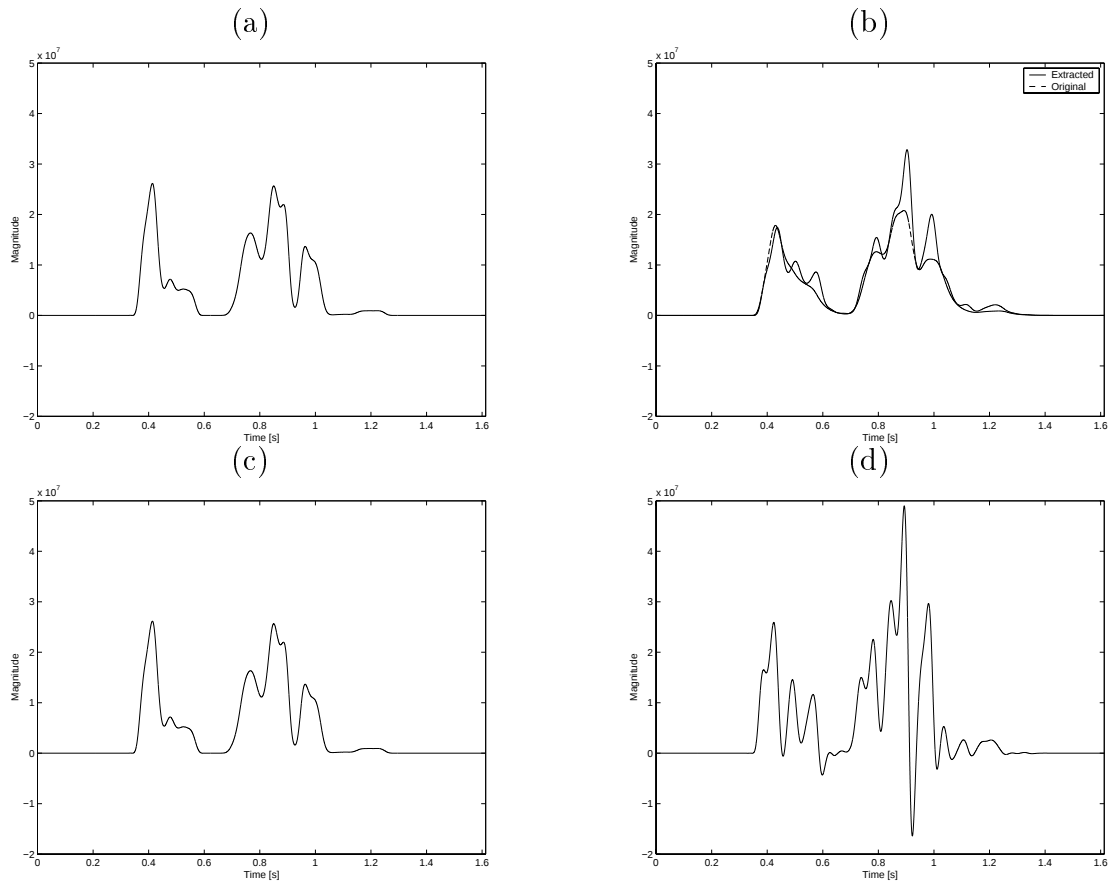


図 2.4: パワーエンベロープ逆フィルタ処理によるパワーエンベロープ回復.(a) 原信号,(b) 残響時間 $T_R = 0.5$ [s] の残響信号,(c) 観測信号のパワーエンベロープを既知とした場合の回復信号,(d) 回復信号

第3章 パワーエンベロープ抽出法

パワーエンベロープ逆フィルタ処理により、残響信号のパワーエンベロープが回復される可能性がある。しかし、その回復精度が残響信号のパワーエンベロープを抽出する精度に依存するという問題点がある。パワーエンベロープ逆フィルタ処理により良好に残響波形のパワーエンベロープを回復して残響を抑圧するためには、パワーエンベロープの抽出精度が高い手法を用いる必要がある。本章では、パワーエンベロープ抽出法として7つ提案する。そのうち、パワーエンベロープの抽出精度が最も高いものを調査する。

3.1 パワーエンベロープ抽出法の提案

残響信号からパワーエンベロープを抽出する手法として以下に示す7つの手法を提案する。

1. ヒルベルト変換による方法
2. 信号のパワーによる方法
3. 包絡線検波による方法
4. 集合平均による方法
5. 集合平均による方法 + ヒルベルト変換
6. 集合平均による方法 + パワー
7. 集合平均による方法 + 包絡線検波

それぞれについて、パワーエンベロープを抽出する原理を以下に述べる。

3.1.1 抽出法の原理

ヒルベルト変換による方法

ヒルベルト変換は、信号 $x(t)$ の位相が 90° ずれた信号 $x_a(t)$ を求める操作を言う。式 (3.1) にヒルベルト変換を示す。

$$x_a(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau \quad (3.1)$$

$x(t)$ と $x_a(t)$ は位相が 90° ずれているため直交する。例えば、 $\cos \omega t$ をヒルベルト変換すると $\sin \omega t$ が求められる。ここで、 $x(t)$ と $x_a(t)$ の2乗和を求めると、直交であるため $x(t)$

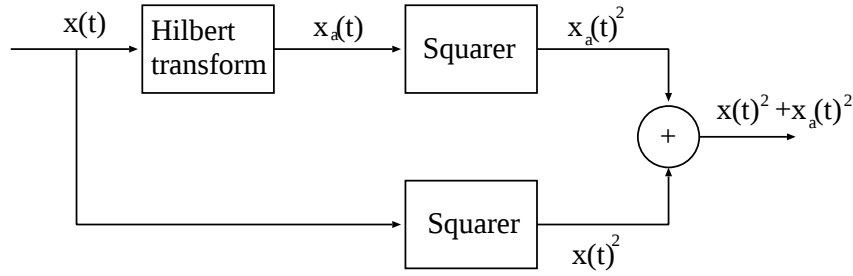


図 3.1: ヒルベルト変換による瞬時振幅の計算

の 2 乗振幅値が求められる。この 2 乗振幅値の時間的に緩やかな変動が $x(t)$ のパワーエンベロープに対応する。図 3.1 にヒルベルト変換により 2 乗振幅を求める簡単なブロック図を示す。

求めた 2 乗振幅から時間的に緩やかな変動であるパワーエンベロープを低域フィルタリングをすることで抽出する。抽出したパワーエンベロープを $\tilde{e}_x(t)^2$ とすると

$$\tilde{e}_x(t)^2 = \text{LPF}[x(t)^2 + x_a(t)^2] \quad (3.2)$$

と表すことができる。ここで、 $\text{LPF}[\cdot]$ は低域フィルタリングを表す。

信号のパワーによる方法

最も単純な方法として、信号のパワーによる方法がある。この手法は、信号のパワーエンベロープは主に低域の周波数成分と考えられることから、信号 2 乗値を求め低域フィルタリングするものである。式 (3.3) にそれを示す。

$$\tilde{e}_x(t)^2 = \text{LPF}[x(t)^2] \quad (3.3)$$

包絡線検波による方法

もう 1 つの方法として積分器を用いた包絡線検波がある。検波後のパワーエンベロープを $x_a(t)^2$ とすると

$$x_a(t)^2 = \exp\left(-\frac{1}{RC}\right) x_a(t-1)^2 + Dx(t)^2 \quad (3.4)$$

$$D = 1 - \exp\left(-\frac{1}{RC}\right) \quad (3.5)$$

で表されるシステムにより求めた。但し、 RC は時定数である。パワーエンベロープを抽出するため、2 乗信号に対して包絡線検波をした。包絡線検波の動作原理は次のようになる。

入力信号の正のサイクルで、積分器は入力信号のピークまで加算していく。ピークを過ぎると時定数に従い積分器の出力は徐々に減衰していく。出力より大きな入力が見られると積分器はさらに加算をする。この繰り返しにより包絡線にほぼ従うことになるが、高周波のリプルが生じる。求めたいパワーエンベロープは時間的に緩やかな変動をするため、低域フィルタリングすることにより高周波のリプルを除去する。

$$\tilde{e}_x(t)^2 = \text{LPF}[x_a(t)^2] \quad (3.6)$$

但し、積分器の時定数を大きくすることにより包絡線に従わなくなることがある。また、時定数を小さくしすぎると大きな高周波のリプルが生じてしまう。このため、時定数の決定には十分注意する必要がある。

集合平均による方法

残響信号のパワーエンベロープは式 (2.17) により求めることができることから、以下のよう考える。

式 (2.11) より白色ガウス過程に従う雑音は無相関であることから、

$$\langle n_1(t), n_2(t) \rangle = 0 \quad (3.7)$$

$$\langle n_k(t), n_m(t) \rangle = 0 \quad (k \neq m) \quad (3.8)$$

$$\langle n_k(t), n_k(t) \rangle = 1 \quad (3.9)$$

となる。ここで、 $\langle, \rangle$ は相関である。今、相異なる白色雑音を掛け合わせても白色雑音であるとする

$$\langle n_k(t), n_m(t) \times n_{k+n}(t) \rangle = 0 \quad (3.10)$$

と考えることができる。式 (3.10) が成立する白色ガウス過程に従う雑音を $\hat{n}(t)$ とすると、観測した残響信号 $y(t)$ が $y(t) = e_y(t)n_k(t)$ としたとき、

$$\hat{y}(t) = y(t)\hat{n}(t) \quad (3.11)$$

$$= e_y(t)n_k(t)\{n_1(t), n_2(t), \dots, n_{k-1}(t), 1, n_{k+1}, \dots\} \quad (3.12)$$

となる。これより、擬似的に $y(t)$ を作ることができる。そして、式 (2.17) の関係から

$$e_y(t)^2 = \langle \hat{y}(t)^2 \rangle \quad (3.13)$$

により、残響信号のパワーエンベロープを求めることができる。

この手法は理論上、無限個の白色雑音を掛け合わせなければならない。しかし、シミュレーションでは不可能であるため有限個の白色雑音を掛け合わせる。このため、求めたパワーエンベロープは計算誤差となる高周波数成分を含んでいる。そこで、求めたパワーエンベロープを低域フィルタリングすることで計算誤差を除去する。

集合平均による方法 + ヒルベルト変換, パワー, 包絡線検波

集合平均による方法と同様の考え方を、ヒルベルト変換, パワー, 包絡線検波による方法に持ち込む。すなわち、残響信号に雑音を掛け合わせた信号から、それぞれの方法でパワーエンベロープを抽出してその集合平均を求める。各手法をそれぞれ $\text{Ext}_1[\cdot]$, $\text{Ext}_2[\cdot]$, $\text{Ext}_3[\cdot]$, とすると

$$\tilde{e}_x(t)^2 = \langle \text{Ext}_k[\{x(t) \times \hat{n}(t)\}^2] \rangle \quad \because k = 1, 2, 3 \quad (3.14)$$

と表される。この手法も、集合平均による方法と同様に掛け合わせる白色雑音が有限個であり計算誤差が生じるため、低域フィルタリングをすることで除去する。

3.1.2 低域通過フィルタ

パワーエンベロープを抽出するとき、前節で示したように全手法において後段で低域フィルタリングをしている。そこで、低域通過フィルタの種類とカットオフ周波数, 特性などを決める必要がある。

金寺ら [10] は、音声認識における変調周波数の重要性について述べている。これによると、最も有用な言語情報が $1 \sim 16$ [Hz] の変調周波数であり、知覚実験の結果とも一致していることを示している。また、音声認識器への貢献度が高いのもこの周波数であることも示している。この結果から、本研究におけるパワーエンベロープの周波数領域は 20 [Hz] までとする。

パワーエンベロープを抽出するとき、その周波数領域以外の周波数はできる限り含まない方がよい。また、重要な周波数成分を残すため通過域はできる限り平坦な方がよいことから、係数感度を考慮して低域通過フィルタは Butterworth 型, 5 次のフィルタを 3 段従属接続したものを使用する。通過域の振幅特性を考慮してカットオフ周波数を決定した結果 25 [Hz] となった。図 3.2 にその振幅特性を示す。

ここで、位相特性に関して考慮していないが振幅の特徴を反映する MTF の理論に基づいているため無視できる。

3.2 抽出法の比較実験

3.2.1 比較実験の条件

パワーエンベロープ抽出法の比較をする。調査には MTF の理論に基づきモデル化した式 (2.8) により原信号を作成する。以下にそれぞれの作成方法について述べる。

1. まず、最も単純なパワーエンベロープを考える。

$$e_x(t)^2 = 1 + \sin 2\pi F t \quad (3.15)$$

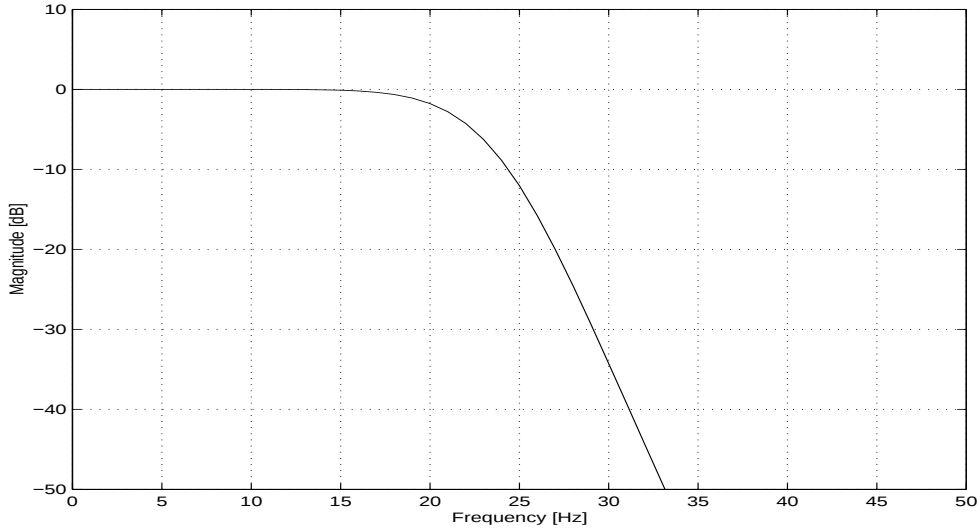


図 3.2: 低域フィルタの振幅特性.

ここで F はパワーエンベロープの周波数であり、 $F = 10$ [Hz] とする。以後、このパワーエンベロープを正弦波のパワーエンベロープと呼ぶ。

2. 音声のパワーエンベロープはいくつかの周波数を含んでいると考えられる。そこで、式 (3.15) をやや複雑にしたものとして、1 ~ 20 [Hz] までの周波数成分を含んだパワーエンベロープを考える。

$$e_x(t)^2 = \sum_{k=1}^{20} \sin(2\pi kt + \hat{\theta}_k) \quad (3.16)$$

但し、 $\hat{\theta}_k$ はランダムな値である。これより 1 ~ 20 [Hz] の周波数成分を含むパワーエンベロープが作成できる。以後、このパワーエンベロープを複合音のパワーエンベロープと呼ぶ。

3. 音声のパワーエンベロープを模擬したものとしてパワーエンベロープをランダム信号から作成することを考える。作成したランダム信号から、カットオフ周波数 20 [Hz] の LPF でフィルタリングすることで求められる。以後、このパワーエンベロープをランダム信号からのパワーエンベロープと呼ぶ。

以上 3 つの変調信号を用いてパワーエンベロープ抽出法の検証を行う。

3.2.2 検証結果

前節で示した変調信号 (原信号とする) に提案したパワーエンベロープ抽出法をそれぞれ施してパワーエンベロープの抽出精度を測定する。ここで、ヒルベルト変換による方法では搬送波の影響により、求めた瞬時振幅と信号のパワーエンベロープとでは大きさが異なる。そこで、非線形最小 2 乗法 [13] によりヒルベルト変換後の信号に掛ける振幅項を求

めた。その結果、振幅項の値は 0.501, 誤差量は 16.17 [dB] である。また、積分器の時定数はパワーエンベロープに含まれる周波数領域を考慮して 10 [ms] とした。集合平均による方法において、白色雑音を掛け合わせる回数は 5000 回とした。

抽出精度の評価尺度には相関値と以下に示す SNR を用いた。

$$\text{SNR} = 10 \log_{10} \frac{\sum_n e_x^2(t)}{\sum_n \{e_x(t) - \tilde{e}_x(t)\}^2} \quad (3.17)$$

ここで、変調信号作成時に求めたパワーエンベロープを $e_x(t)^2$, 抽出したパワーエンベロープを $\tilde{e}_x(t)^2$ とする。相関値, SNR を用いた理由は、パワーエンベロープの類似性と絶対誤差を測定するためである。相関値が限りなく 1 に近づき、SNR が大きな値になれば残響信号から正確にパワーエンベロープが抽出できたことになる。

検証シミュレーションをするにあたり、変調信号の搬送波である白色雑音の実現値が異なることで、パワーエンベロープの抽出精度が変化することを考慮にいれなければならない。本シミュレーションでは、搬送波である白色雑音の実現値を 100 種類用いた。図 3.3(a) に相関値, (b) に SNR の平均値と標準偏差を示す。横軸は原信号パワーエンベロープの種類であり、エラーバーはその標準偏差を示している。棒グラフは提案したパワーエンベロープ抽出法であり、左から (1) ヒルベルト変換, (2) パワー, (3) 包絡線検波, (4) 集合平均, (5) 集合平均+ヒルベルト変換, (6) 集合平均+パワー, (7) 集合平均+包絡線検波による手法である。また、原信号においてパワーエンベロープの抽出精度が比較的良好な手法は、式 (2.7) により残響を付加して相関値と SNR を求める。用いたインパルス応答の残響時間は 0.1, 0.3, 0.5, 1.0, 2.0 [s] である。結果を図 3.4, 3.5 に示す。実線, 破線, 点破線は、原信号のパワーエンベロープがそれぞれ正弦波のパワーエンベロープ, 複合音のパワーエンベロープ, ランダム信号からのパワーエンベロープの場合を示している。

3.2.3 考察

パワーエンベロープの抽出精度は逆フィルタ処理による回復精度に影響を与える。そのため、パワーエンベロープ抽出法を評価する相関値と SNR は大きな値が望ましい。図 3.3 より、ヒルベルト変換, パワー, 集合平均, 集合平均+パワーによる方法は、相関値, SNR とともに大きく抽出精度が比較的良好ことがわかる。またこれらの手法は、図 3.4, 3.5 からわかるように、残響を付加してもパワーエンベロープの抽出精度が低下しない。そのため、ヒルベルト変換, パワー, 集合平均, 集合平均+パワーによる方法は、パワーエンベロープの抽出法として同レベルであると言える。本研究では、この中で最も標準偏差が小さいヒルベルト変換による方法でパワーエンベロープを抽出することにする。また、ヒルベルト変換による方法は SNR が 20 [dB] を超える程度が限界であると考えられる。

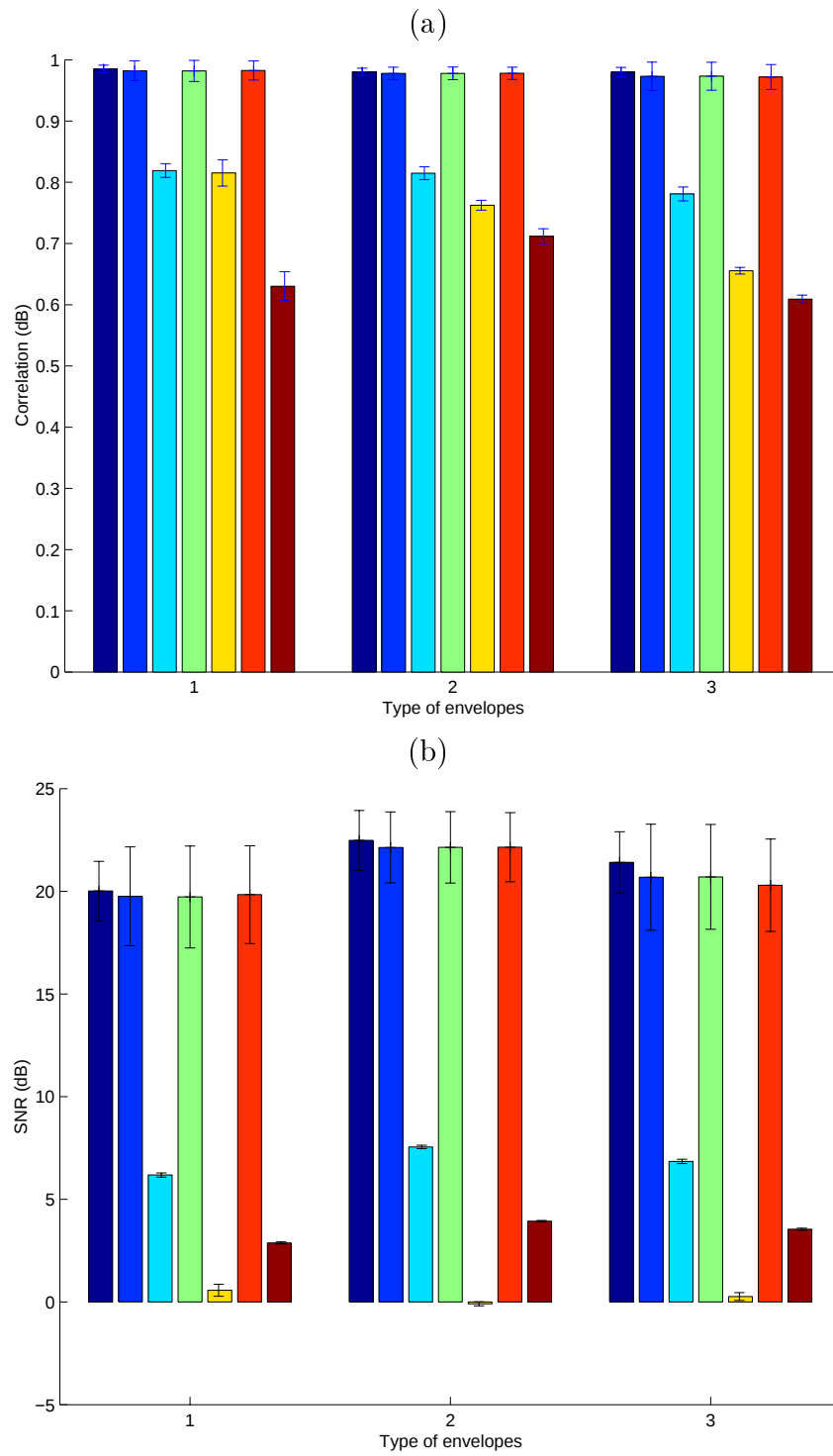


図 3.3: パワーエンベロープの抽出精度.(a) 相関値の平均値と標準偏差, (b) SNR の平均値と標準偏差

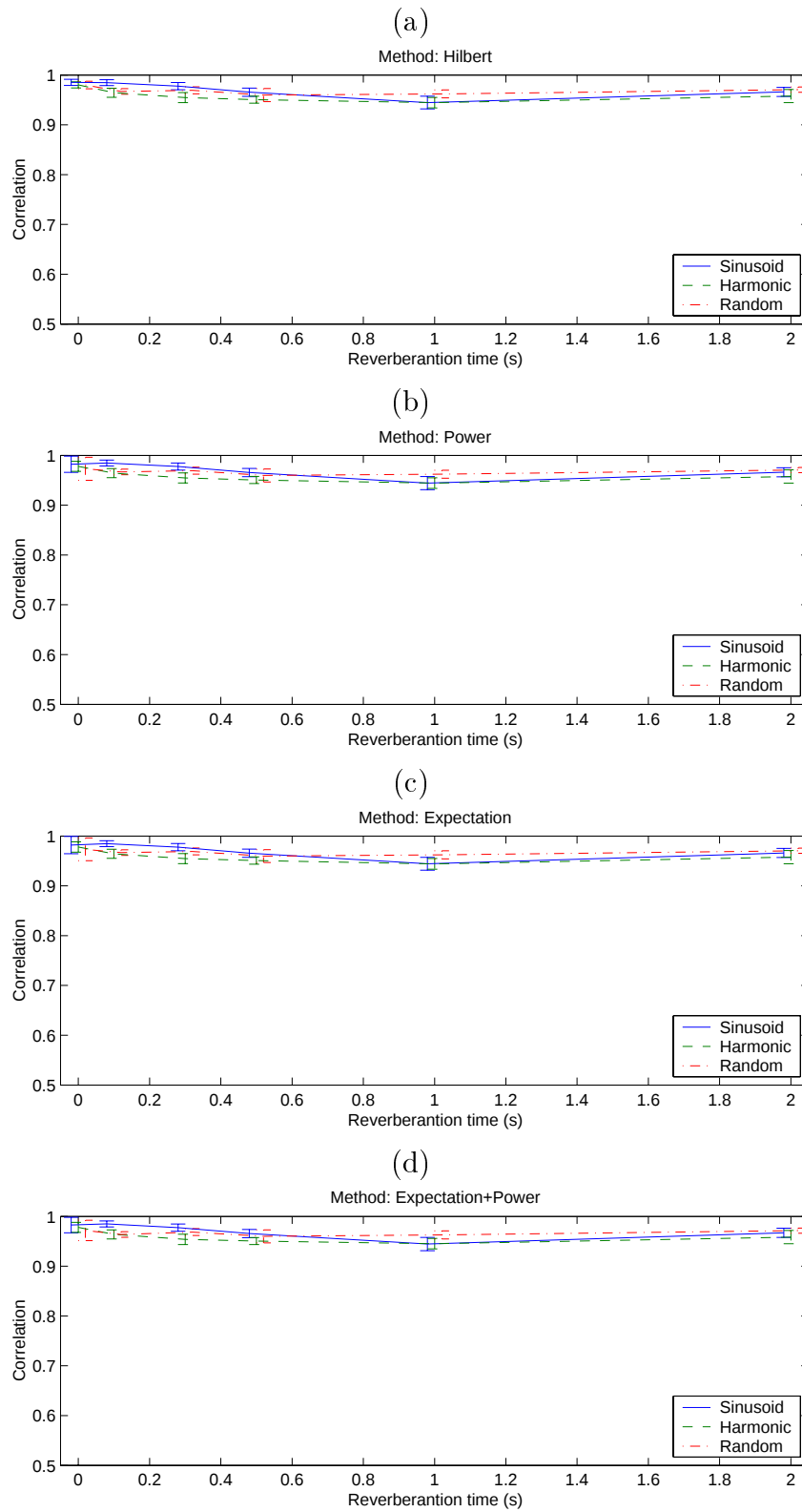


図 3.4: 残響を付加したときの相関値の関係.(a) ヒルベルト変換による方法,(b) パワーによる方法,(c) 集合平均による方法,(d) 集合平均+パワーによる方法

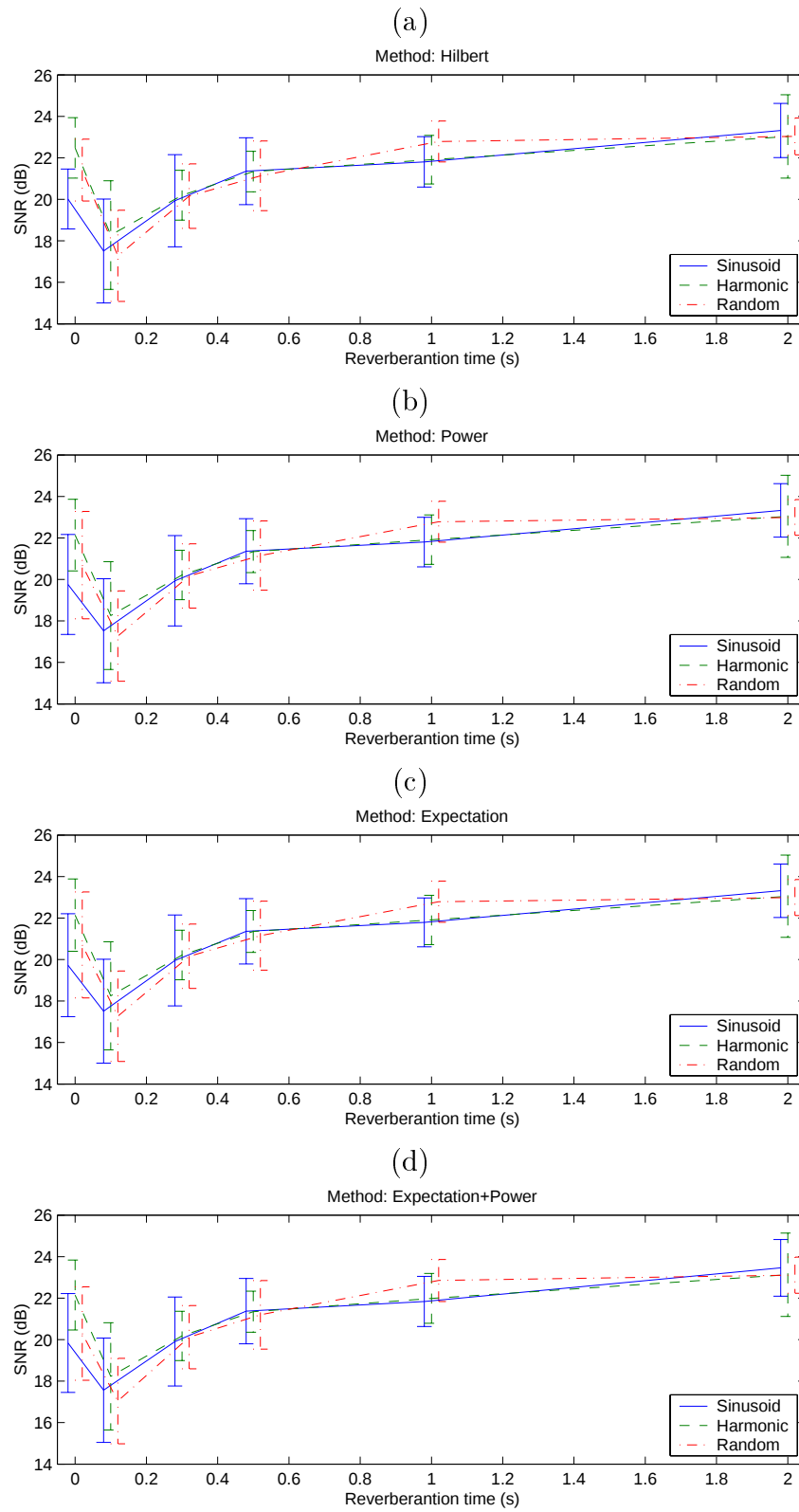


図 3.5: 残響を付加したときの SNR の関係.(a) ヒルベルト変換による方法, (b) パワーによる方法,(c) 集合平均による方法,(d) 集合平均+パワーによる方法

第4章 残響時間の推定

広林らが提案しているパワーエンベロープ逆フィルタ処理は、残響信号のパワーエンベロープを回復するものである。パワーエンベロープ逆フィルタ処理は、伝達系の測定なしに残響信号のパワーエンベロープを回復するが残響時間が既知でなければならない。本章では、残響時間を推定する手法として3つ提案する。そのうち、推定法として最適な手法を調査する。

4.1 変調度に関する仮定

残響時間の推定を行うにあたり、変調度に関して以下のような仮定をする。

- 原信号の変調度は1である

パワーエンベロープのひずみが室内の残響音による影響である場合、残響信号の変調度 m_y は式 (2.4) より求めることができる。このとき変調度 m_y は $0 \leq m_y \leq 1$ であり、原信号の変調度に対する相対量である。そのため、原信号の変調度は既知でなければならない。また、音声において、無音区間ではパワーがないと考えられることから、原信号の変調度を1と仮定する。

4.2 インパルス応答の定数に関する仮定

式 (2.10) において残響時間の他にインパルス応答の増幅度を決定する定数 a を決定しなければならない。そこで残響が増幅よりも主に位相遅れに強く関係すると考え、以下のような仮定をした。

- インパルス応答による増幅度は1である

本研究ではインパルス応答における包絡線の2乗和が1になるように a を決定した。

4.3 パワーエンベロープ逆フィルタの特徴

式 (2.20) にパワーエンベロープ逆フィルタを示した。その振幅特性を図 4.1 に示す。

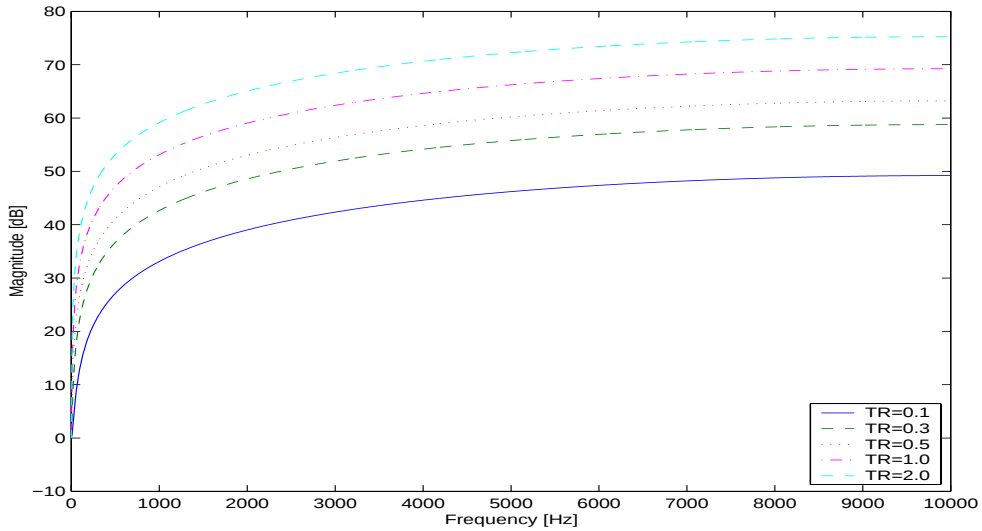


図 4.1: パワーエンベロープ逆フィルタの特徴.

図 4.1 より、高周波数成分が強調されていることがわかる。特に回復に用いる残響時間が長ければ長いほど、より高周波数成分が強調される傾向にある。残響時間の推定法を提案するにあたり、この特徴を用いることを考える。

4.4 残響時間の推定方法

4.4.1 残響時間推定法の提案

残響時間の推定法として以下に挙げる 3 つの手法を提案する。

1. 負の面積による方法
2. 変調伝達関数による方法
3. 高周波数成分による方法

次にこれらについて詳しく述べる

4.4.2 推定法の原理

負の面積による方法

残響は、時間的に正方向にパワーエンベロープのピークの山をならし、ならされたものが谷を埋める。パワーエンベロープ逆フィルタは、図 4.1 から明らかなように高周波数成分を強調することでこのならされたものを削り取り時間的に逆方向に谷を強調して山を盛り上げる。もし、推定した残響時間 $\hat{T}_R$ が付加した残響時間 T_R よりも大きいとき、谷

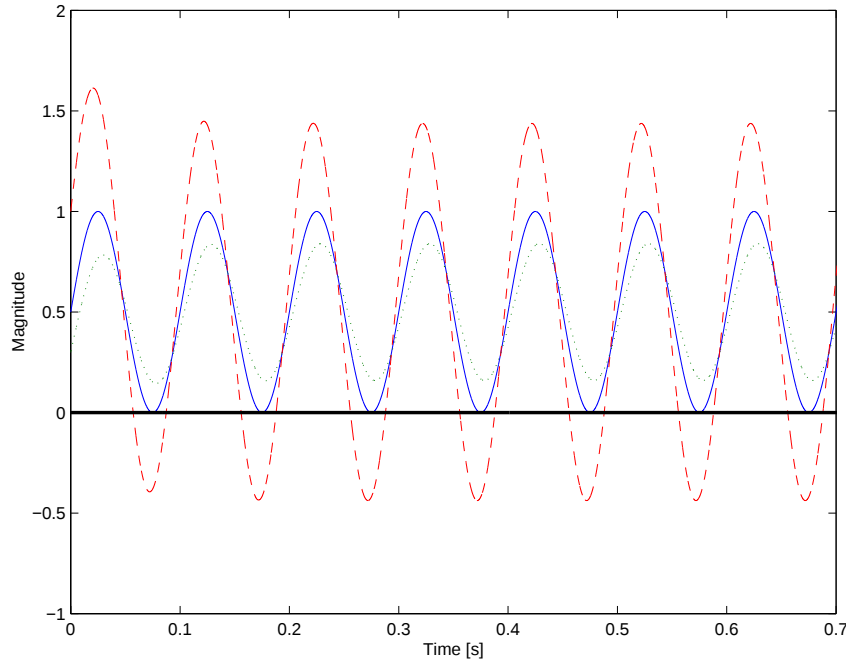


図 4.2: 過剰な回復により現れる負の面積 (付加した残響時間 $T_R = 0.5$ [s]). 実線: 正確な回復のパワーエンベロープ ($\hat{T}_R = 0.5$), 点線: 回復不足のパワーエンベロープ ($\hat{T}_R = 0.3$), 破線: 過剰回復のパワーエンベロープ ($\hat{T}_R = 1.0$)

が必要以上に強調されることでパワーエンベロープが負の値になることがある (図 4.2 参照)。しかし、信号の 2 乗包絡線であるパワーエンベロープにおいて、負の値を持つことがないことは明らかである。これは、回復パワーエンベロープ $\hat{e}_x(t)^2$ を用いて

$$\sum_{t=0}^T \min\{\hat{e}_x(t)^2, 0\} = 0 \quad (4.1)$$

と表すことができる。ここで $\min(\cdot, 0)$ は 0 以下の値のみ取り出すことを示す。また、 T は信号の長さである。

本研究において、原信号の変調度は 1 と仮定していることから、逆フィルタ処理により回復したパワーエンベロープが負の値をもつ直前の残響時間を求めることで、観測信号の情報のみから残響時間が自動推定できると考えられる。

変調伝達関数による方法

Houtgast ら [7] は残響環境下における変調伝達関数を式 (2.4) のように定義している。式 (2.4) より、変調周波数における残響信号の変調度がわかれば残響時間 T_R が式 (4.2) で計算できる。

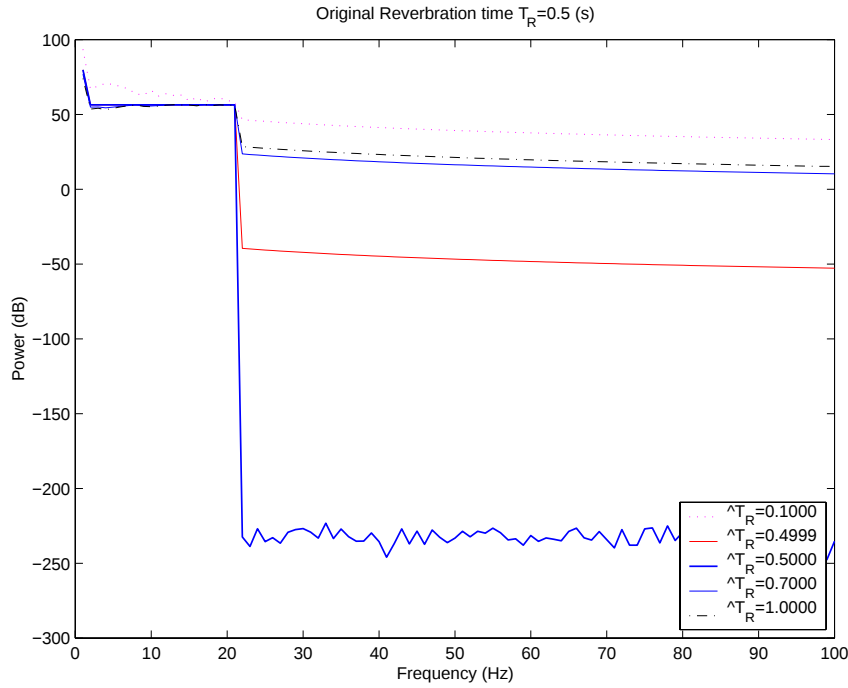


図 4.3: 回復時に用いる残響時間と回復パワーエンベロープにおける周波数成分の関係.

$$\hat{T}_R = \frac{13.8}{\omega} \sqrt{\frac{1}{m_y} - 1} \quad (4.2)$$

ここで ω は変調周波数であり、ここでは 10 [Hz] とした。 m_y は残響信号におけるパワーエンベロープの変調度であり $0 \leq m_y \leq 1$ である。残響信号におけるパワーエンベロープの変調度は原信号における変調度との相対値であることから、残響信号の局所最大点と局所最小点の距離を L_y 、原信号の局所最大点と局所最小点の距離を L_x とすると

$$m_y = \frac{L_y}{L_x} \quad (4.3)$$

と表すことができる。本研究において原信号の変調度は 1 と仮定しているため、残響信号のパワーエンベロープを正規化した時の局所最大点と局所最小点の距離 $L_y(\omega)$ が変調度と見なすことができる。これより式 (4.2) で残響時間を求めることができる。

高周波数成分による方法

式 (2.17) により作成した残響信号パワーエンベロープを、様々な残響時間により回復する。回復パワーエンベロープ $\hat{e}_x(t)^2$ の周波数成分を調査した結果、図 4.3 が得られた。ここで、付加した残響時間は $T_R = 0.5$ [s] である。

図 4.3 より、回復に用いる残響時間 $\hat{T}_R$ が正しいとき、すなわち $\hat{T}_R = T_R$ のとき高周波数成分の強調がないことがわかる。そして、 $\hat{T}_R \neq T_R$ のとき高周波数成分は残存している。これより、 $\hat{T}_R = T_R$ のとき回復したパワーエンベロープの高周波数成分が最も少ないと考えられる。様々な残響時間で回復したパワーエンベロープの周波数成分を $\hat{F}_x(\omega, \hat{T}_R)$ とすると

$$\hat{T}_R = \min \left\{ \sum_{\omega=20}^{f_s/2} \hat{F}_x(\omega, \hat{T}_R) \right\} \quad (4.4)$$

となる $\hat{T}_R$ を求めれば良いことになる。ここで f_s はサンプリング周波数である。

4.5 残響時間推定の比較実験

4.5.1 検証シミュレーション

残響時間推定法の検証を行う。原信号は、パワーエンベロープ抽出法と同様に正弦波のパワーエンベロープ、複合音のパワーエンベロープ、ランダム信号からのパワーエンベロープを用いる。これら 3 種類の原信号に残響を付加して、残響時間推定の検証を行う。付加する残響時間は $T_R = 0.1, 0.3, 0.5, 1.0, 2.0$ [s] の 5 種類を用いた。

4.5.2 検証結果

原信号に残響インパルス応答を畳み込むことで残響信号を作成する。その後、ヒルベルト変換によりパワーエンベロープを抽出して、前節で示した手法により残響時間を推定する。残響時間推定シミュレーションには白色雑音の実現値を 100 種類用い、その平均値と標準偏差を求めた。負の面積、変調伝達関数、高周波数成分による方法により推定した残響時間の結果を図 4.4, 4.5, 4.6 にそれぞれ示す。(a) は残響信号パワーエンベロープの理論値を用いて推定した残響時間、(b) は残響信号からヒルベルト変換を用いて抽出したパワーエンベロープを用いて推定した残響時間である。実線、破線、点破線は、それぞれ原信号のパワーエンベロープが正弦波のパワーエンベロープ、複合音のパワーエンベロープ、ランダム信号からのパワーエンベロープの場合を示している。

4.5.3 考察

図の点線は $\hat{T}_R = T_R$ を示しており、正確な残響時間の推定が行われればグラフは点線に従う。そして図 4.4, 4.5, 4.6 の (a) は、MTF の理論に従う残響信号のパワーエンベロープを用いているため、正確に残響時間が推定できなければならない。

図 4.4 の (a) では正確に残響時間を推定できるが、図 4.5, 4.6 の (a) では正確に残響時間を推定できない。MTF の理論では、パワーエンベロープが単一の周波数成分を持つとき、

変調度と残響時間の関係式 (2.4) が定義できる。しかし、パワーエンベロープが複数の周波数成分を含むとき、式 (2.4) を定義できない。また、変調度を決定するときに求める残響信号の局所最大点と局所最小点の距離に依存しており、その距離を自動的に求めることが困難であることから、変調周波数による方法は正確に残響時間を推定できないと考えられる。高周波数成分による方法では、白色雑音から求めたパワーエンベロープを用いた場合、残響時間が正しく推定できない。また、抽出したパワーエンベロープに対しても推定できない。これは、パワーエンベロープの周波数成分がある帯域内のみであり、MTF の理論に従う残響信号のパワーエンベロープを正確に抽出できないと、残響時間の推定が正確に行えないことを示している。実装上、このような条件を満たすことはできないため、高周波数成分による残響時間の推定は困難であると考えられる。

これに対して負の面積による方法は、MTF の理論に従えば正確に残響時間が推定できる。抽出したパワーエンベロープでは、付加した残響時間が長くなると推定精度が低下する。ここで、抽出した残響信号のパワーエンベロープを残響時間を変数として回復処理をしたパワーエンベロープ $\hat{e}_x(t)^2$ と原信号のパワーエンベロープ $e_x(t)^2$ との相関値と SNR を図 4.7 に示す。(a) ~ (e) はそれぞれ残響時間 $T_R = 0.1, 0.3, 0.5, 1.0, 2.0$ [s] であり、マークしている部分は推定した残響時間である。これより、負の面積による残響時間の推定は抽出したパワーエンベロープの相関値と SNR が比較的高い残響時間を推定していることがわかる。第 2 章でパワーエンベロープ逆フィルタを検証したとき、残響時間を既知とした場合でパワーエンベロープの回復度は、変調信号で 1.77 [dB]、音声で 0.41 [dB] であった。これらの信号から推定した残響時間は各々 0.37 [s], 0.215 [s] であり、パワーエンベロープ回復度は 3.31 [dB], 1.15 [dB] と向上した。これらのことから、抽出した残響信号パワーエンベロープの回復が高い残響時間を推定することになるため、負の面積による残響時間の推定は有効であると考えられる。

残響時間推定誤差は、抽出したパワーエンベロープにおける 20 ~ 30 [Hz] 付近の周波数成分が原因である。本研究ではパワーエンベロープの周波数領域を 20 [Hz] までとしている。それ以外の周波数成分が存在すると、パワーエンベロープ逆フィルタにより強調され誤差の原因となる。このためパワーエンベロープ抽出に用いる低域フィルタは急峻な遮断特性をもつものを使用しているが、図 3.2 から 20 ~ 30 [Hz] 付近の周波数成分が他の周波数と比べて比較的多く残るため、逆フィルタにより強調されてしまい残響時間の推定誤差となる。図 4.1 からわかるように、逆フィルタ処理に用いる残響時間が長いと高周波数成分をより強調するため、付加した残響時間が長くなると推定誤差も大きくなり、精度が低下すると考えられる。この問題は、誤差の原因である周波数成分だけを減衰させることで改善すると考えられる。

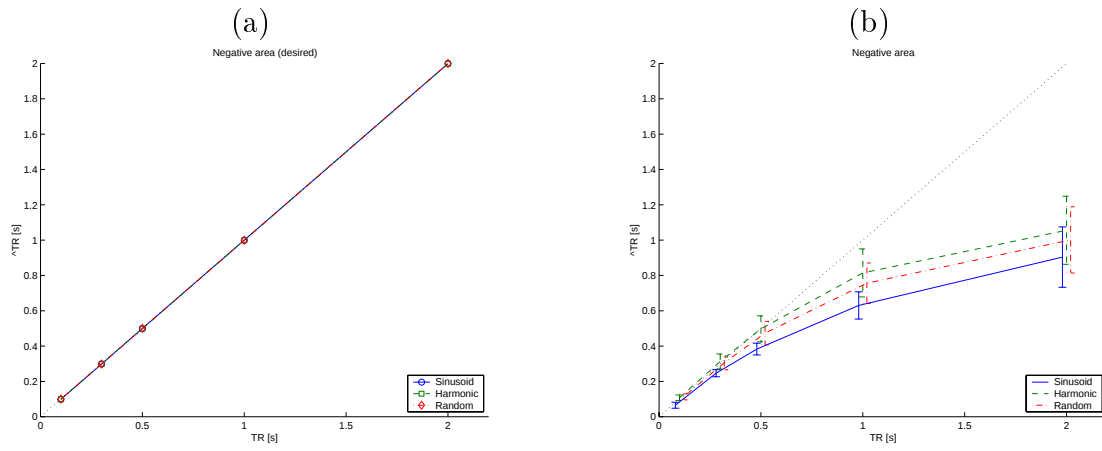


図 4.4: 負の面積による残響時間の推定.

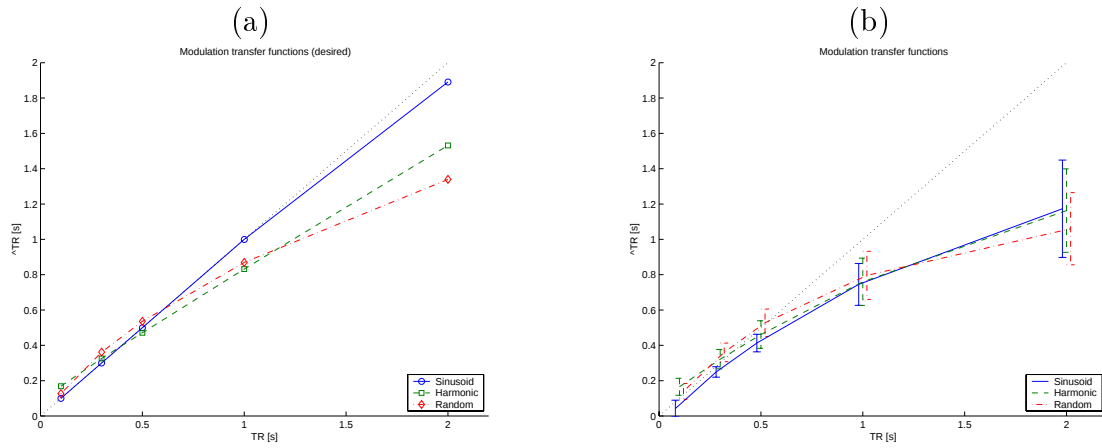


図 4.5: 変調伝達関数による残響時間の推定.

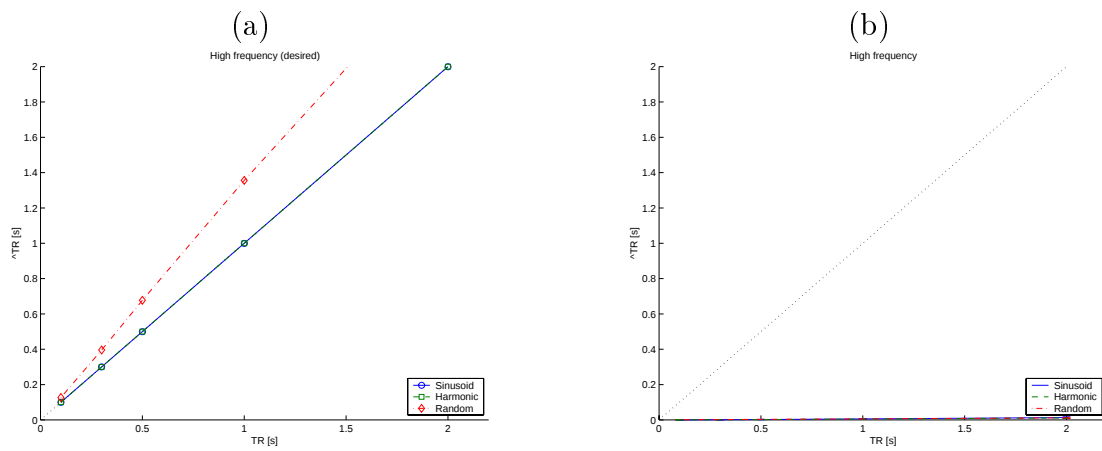
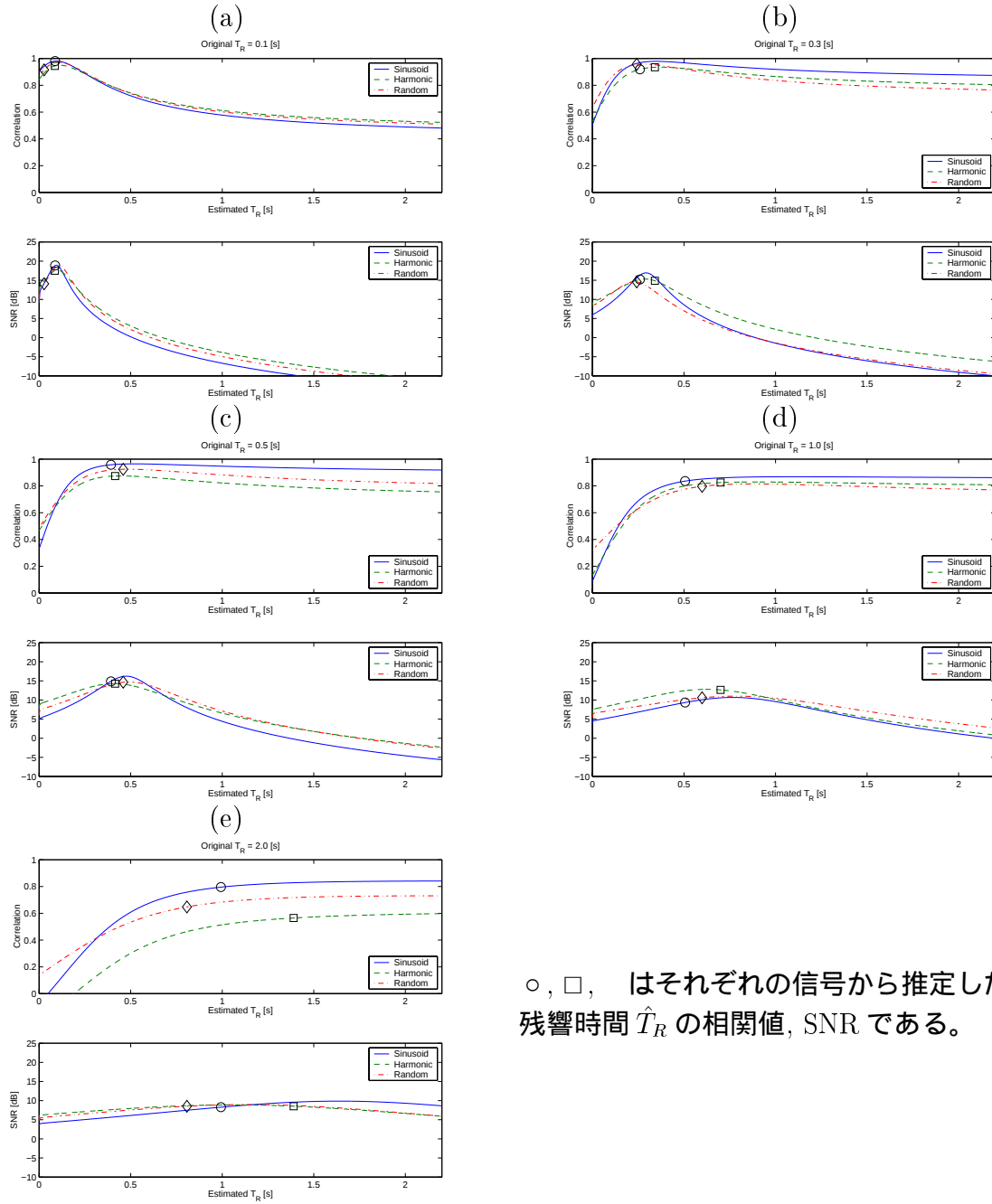


図 4.6: 高周波数成分による残響時間の推定.



○, □, ◇ はそれぞれの信号から推定した
残響時間 $\hat{T}_R$ の相関値, SNR である。

図 4.7: 推定残響時間と相関, SNR の関係. (a) $T_R = 0.1$ [s], (b) $T_R = 0.3$ [s], (c) $T_R = 0.5$ [s], (d) $T_R = 1.0$ [s], (e) $T_R = 2.0$ [s]. (上) 相関値, (下) SNR.

第5章 音声への適用

前章までは、MTF の理論に基づいた原信号を用いてパワーエンベロープ逆フィルタ処理による残響信号パワーエンベロープ回復の原理上の問題点について検証した。本章では、パワーエンベロープ逆フィルタ処理を音声に適用する。そして、残響音声パワーエンベロープを回復する手法としての有効性を検証する。

5.1 音声を用いる問題点

MTF の理論に基づいた原信号は、振幅エンベロープと搬送波である白色雑音との積で表現される。音声を MTF の理論に基づいて振幅エンベロープと白色雑音との積で表現しようとするとき、以下に示す問題点がわかる。

1. 音声において搬送波は調波複合音を含んだものである
2. 音声の振幅スペクトルは共変調ではない

次に、これらについて詳しく述べる。

5.1.1 音声の搬送波

広林らは MTF の理論に基づき、原音声を振幅エンベロープと白色雑音でモデル化している。しかし、実際の音声において搬送波は調波複合音を含んだものであると考えられる。これまで示してきたパワーエンベロープ抽出法や残響時間の推定法は、搬送波を白色雑音として検証してきたため、搬送波を調波複合音としても有効であるかの検証がされていない。そこで、搬送波を調波複合音としてパワーエンベロープの抽出法と残響時間の推定に関して再検証を行う。

5.1.2 音声と共変調

パワーエンベロープ逆フィルタ処理は、残響により歪んだパワーエンベロープを回復する。このとき、広帯域の残響信号に対してパワーエンベロープ逆フィルタ処理により包絡線情報を回復した場合（以後、一括処理と略す）、全ての帯域において包絡線情報が等しくなる。一括処理によりパワーエンベロープを回復した音声のスペクトログラムを図 5.1 に

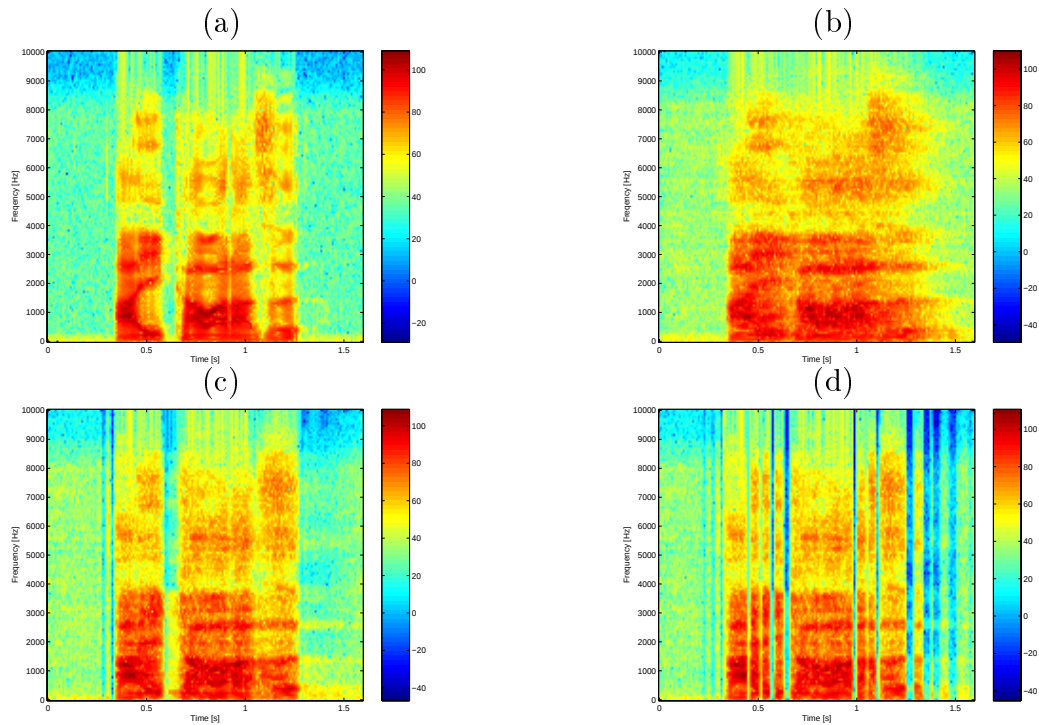


図 5.1: 一括処理における音声のスペクトログラム.(a) 原音声,(b) 残響時間 $T_R = 0.5$ [s] の残響を付加した音声,(c) 観測した音声のパワーエンベロープを既知としたときの回復音声,(d) 残響時間を既知とした回復音声

示す。(a) は原音声,(b) は原音声に残響時間 $T_R = 0.5$ [s] の残響を付加した残響音声, (c) は観測音声のパワーエンベロープを既知とした場合の回復音声,(d) は、残響時間を既知として一括処理によりパワーエンベロープを回復した音声である。図 5.1 の (c),(d) より、一括処理によりパワーエンベロープを回復した場合、包絡線情報は全ての帯域で等しくなることが確認できる。つまり、パワーエンベロープ逆フィルタ処理は回復処理をする帯域内で信号が共変調しているとして包絡線情報を回復することがわかる。一方、(a) の原音声のスペクトログラムから、各帯域において包絡線情報の相違が確認できる。これより、一括処理により音声のパワーエンベロープを回復しても、各帯域の包絡線情報を正確に回復できないことがわかる。従って、パワーエンベロープ逆フィルタ処理を残響音声に適用する場合、帯域分割を行う必要があると考えられる。

5.2 調波複合音の搬送波

音声にパワーエンベロープ逆フィルタ処理を適用する場合、音声の搬送波が調波複合音を含んでいることが問題となる。これまで示してきたパワーエンベロープ抽出法と残響時間推定法は、MTF の理論に基づいた信号で検証をしてきたため搬送波を調波複合音にしても有効に機能するかどうか不明である。そこで、搬送波を調波複合音にしてパワーエン

ベロープ抽出法と残響時間推定法の検証を行う。

5.2.1 検証シミュレーション

搬送波を調波複合音として、パワーエンベロープ抽出法と残響時間推定法の検証を行う。パワーエンベロープの抽出ではヒルベルト変換による方法について、残響時間の推定では負の面積による方法について検証を行う。

5.2.2 検証結果

まず、パワーエンベロープ抽出法の検証を行う。パワーエンベロープ抽出に用いる手法はヒルベルト変換による方法である。3.2 節の検証と同様に、原信号のパワーエンベロープは (1) 正弦波のパワーエンベロープ, (2) 複合音のパワーエンベロープ, (3) ランダム信号からのパワーエンベロープの 3 種類を用い、搬送波は調波複合音とした。調波複合音 $n_c(t)$ は

$$n_c(t) = \sum_{k=1}^N \sin(2\pi F_k t) \quad (5.1)$$

と表すことができる。ここで、 $F_1 = 100$ [Hz] とし、サンプリング周波数が 20 [kHz] であり、全周波数帯域に成分を持たせるため $N = 100$ とした。そして、原信号 $x(t)$ は振幅エンベロープ $e_x(t)$ を用いて

$$x(t) = e_x(t)n_c(t) \quad (5.2)$$

と表すことができる。

原信号からのパワーエンベロープ抽出精度を図 5.2 に示す。(a) は相関値, (b) は SNR を示している。横軸は原信号パワーエンベロープの種類であり、エラーバーはその標準偏差を示している。また、残響が付加された信号に対してパワーエンベロープ抽出した結果を図 5.3 に示す。実線, 破線, 点破線は、それぞれ原信号のパワーエンベロープが正弦波のパワーエンベロープ, 複合音のパワーエンベロープ, ランダム信号からのパワーエンベロープの場合を示している。

次に、残響時間推定法について検証を行う。4.5 節の検証と同様に、原信号のパワーエンベロープは (1) 正弦波のパワーエンベロープ, (2) 複合音のパワーエンベロープ, (3) ランダム信号からのパワーエンベロープの 3 種類を用い、搬送波は調波複合音とする。調波複合音は式 (5.1) により求め、原信号は式 (5.2) により求めた。結果を図 5.4 に示す。(a) は残響信号パワーエンベロープの理論値の理論値を用いて推定した残響時間、(b) は残響信号からヒルベルト変換を用いて抽出したパワーエンベロープを用いて推定した残響時間である。また、残響時間を変数として回復処理をしたパワーエンベロープ $\hat{e}_x(t)^2$ と原信号のパワーエンベロープ $e_x(t)^2$ との相関値と SNR を図 5.5 に示す。(a) ~ (e) はそれぞれ残響時間 $T_R = 0.1, 0.3, 0.5, 1.0, 2.0$ [s] であり、マークしている部分は推定した残響時間である。

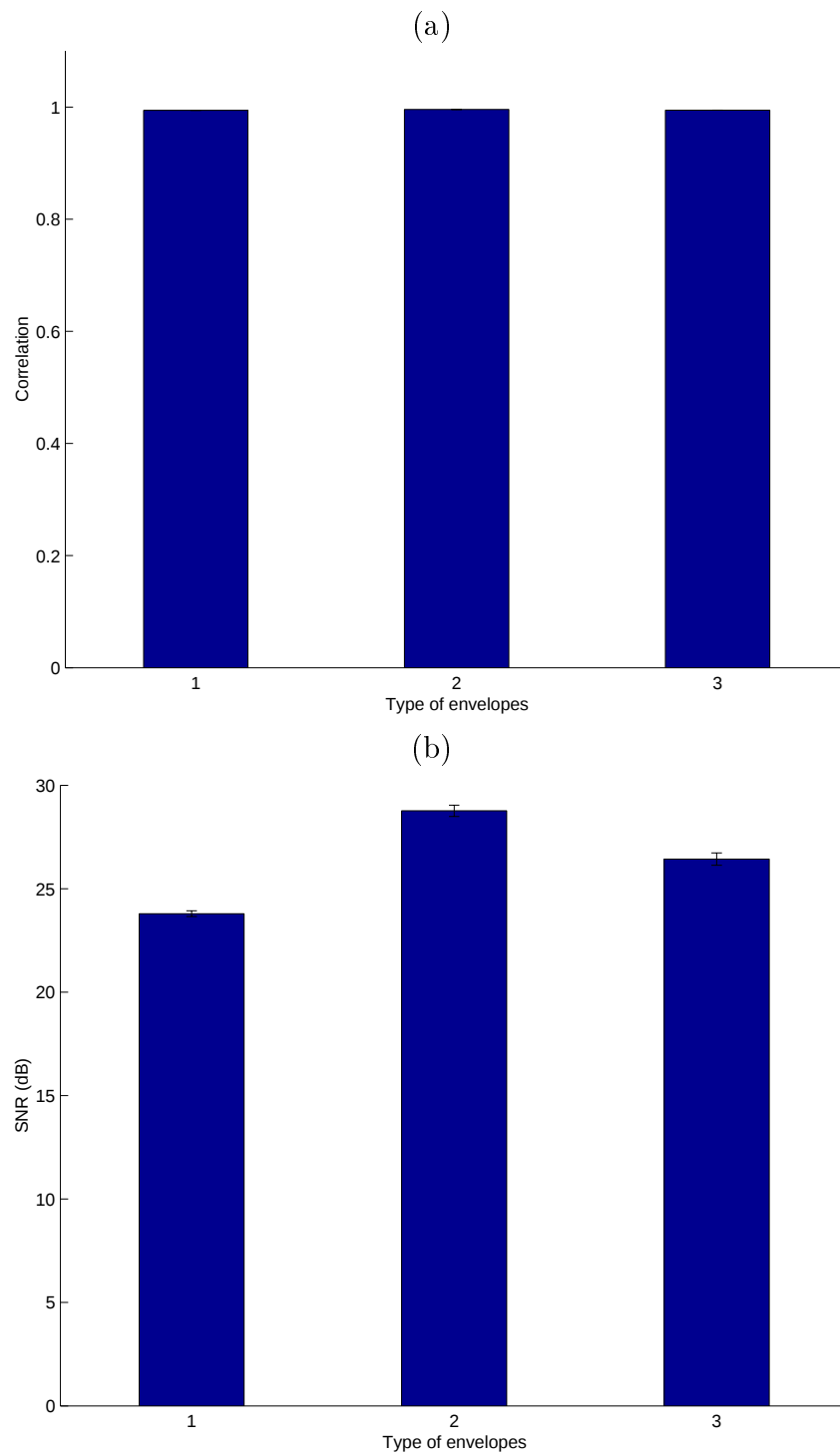


図 5.2: 搬送波が調波複合音のときのパワーエンベロープの抽出精度.

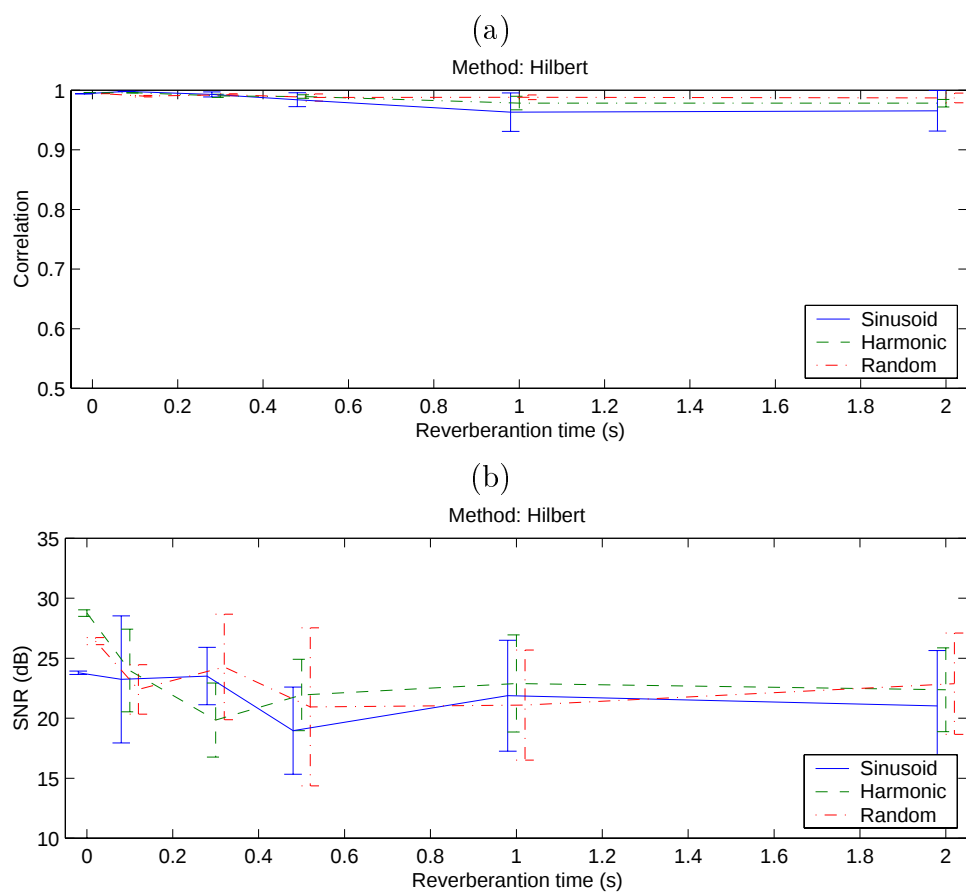


図 5.3: 残響を付加したときの相関値と SNR の関係.(a) 相関値,(b)SNR

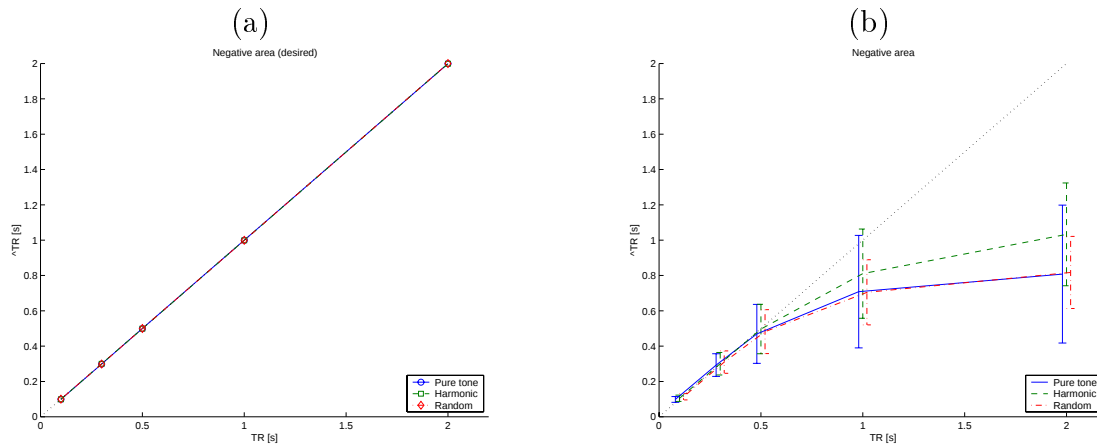


図 5.4: 搬送波が調波複合音の場合の負の面積による残響時間の推定.(a) 残響信号パワーエンベロープの理論値を用いて推定した残響時間,(b) 残響信号からヒルベルト変換を用いて抽出したパワーエンベロープを用いて推定した残響時間

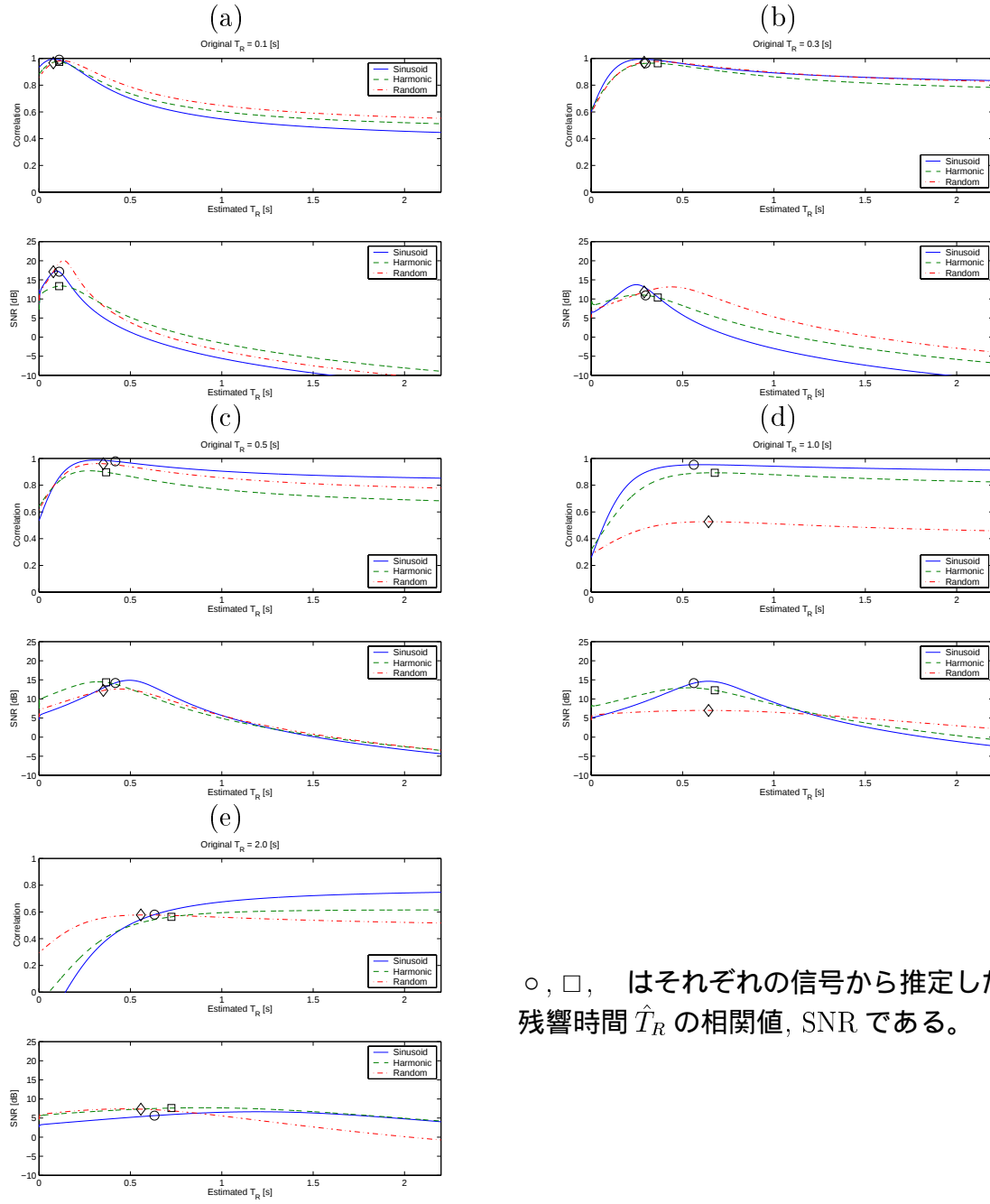
図 5.4,5.5 において、実線、破線、点破線は、それぞれ原信号のパワーエンベロープが正弦波のパワーエンベロープ、複合音のパワーエンベロープ、ランダム信号からのパワーエンベロープの場合を示している。

5.2.3 考察

図 5.2 より、原信号パワーエンベロープの種類が異なっても相関値は 1 に近く SNR は 25 [dB] 以上の値を示している。図 5.3 より、残響が付加された場合のパワーエンベロープの抽出精度は大きく低下することはない。また、搬送波が白色雑音の場合である図 3.3 と図 3.4(a),3.5(a) と比較しても相関値と SNR が低下していない。以上より、搬送波を調波複合音としてもパワーエンベロープ抽出法には影響がないと考えられる。

図 5.4(a) より、MTF の理論に従う残響信号パワーエンベロープの場合、正確に残響時間を推定できる。抽出したパワーエンベロープでは、付加した残響時間が長くなると抽出精度が低下する。しかし推定した残響時間を用いて回復処理を施した図 5.5 より、パワーエンベロープ回復度の高い残響時間を推定していることから、搬送波を調波複合音にしても残響時間の推定には影響がないと考えられる。

以上より、搬送波を調波複合音としてもパワーエンベロープ抽出法と残響時間推定法は有効であることが確認された。



○, □, ◇ はそれぞれの信号から推定した残響時間 $\hat{T}_R$ の相関値, SNR である。

図 5.5: 推定残響時間と相関, SNR の関係. (a) $T_R = 0.1$ [s], (b) $T_R = 0.3$ [s], (c) $T_R = 0.5$ [s], (d) $T_R = 1.0$ [s], (e) $T_R = 2.0$ [s]

5.3 非共変調信号の回復

前章まで用いてきた振幅変調信号は共変調であった。パワーエンベロープ逆フィルタ処理は観測信号が共変調信号として回復されるため、前章まで用いてきた振幅変調信号のパワーエンベロープは回復できるが、非共変調信号の包絡線情報は回復できないと考えられる。本節では、非共変調信号を作成して帯域分割による残響信号パワーエンベロープの回復について検証する。

5.3.1 検証シミュレーション

原信号として非共変調信号を作成する。搬送波は白色雑音とし、サンプリング周波数は 20 [kHz] とした。まず、白色雑音をフィルタバンクで帯域分割する。ここで、分割方法は 10 [kHz] までは 0 ~ 2.5 [kHz], 2.5 ~ 5 [kHz], 5 ~ 7.5 [kHz], 7.5 ~ 10 [kHz] の 4 等分割とした。分割したそれぞれの白色雑音に振幅エンベロープ $e_x(t)$ を掛け合わせる。 $e_x(t)$ は

$$e_x(t) = \sqrt{1 + \sin 2\pi F_k t} \quad (5.3)$$

により求めた。 F_k は周波数の低い方から 4, 6, 10, 16 [Hz] とした。この時、原信号の変調度が 1 となるように、全ての変調信号が 0.5 [s] で 0 となるようにした。最後に各帯域ごとに作成した信号を合成することで、原信号を作成することができる。この原信号に残響時間 $T_R = 0.5$ [s] の残響インパルス応答を畳み込み残響信号を作成する。この残響信号をパワーエンベロープ逆フィルタ処理で回復し、その回復度を評価する。

5.3.2 検証結果

フィルタバンクにより帯域分割をした各信号に対して、パワーエンベロープ逆フィルタ処理を施すことによりパワーエンベロープを回復する。比較のため、一括処理によりパワーエンベロープの回復も行う。

図 5.6 に時間波形を、図 5.7 にスペクトログラムをそれぞれ示す。(a) は原信号、(b) は残響時間 $T_R = 0.5$ [s] の残響信号、(c) は各帯域における観測信号のパワーエンベロープを既知とした回復信号、(d) は帯域分割による回復信号、(e) は一括処理における観測信号のパワーエンベロープを既知とした回復信号、(f) は一括処理における回復信号である。

図 5.6 より、時間波形では分割処理による回復と一括処理による回復との差がほとんどないように見える。図 5.7(a) において各帯域の包絡線情報が異なっており、(b) では残響によりその包絡線情報が歪んでいる。(c) から各帯域における観測信号のパワーエンベロープが既知であれば包絡線情報が良く回復されていることがわかる。(d) から、帯域分割することによる各帯域でのパワーエンベロープ回復効果が見られる。これに対して一括処理をした (e)、(f) では、全ての帯域において包絡線情報が等しくなり包絡線情報が回復されていないことがわかる。

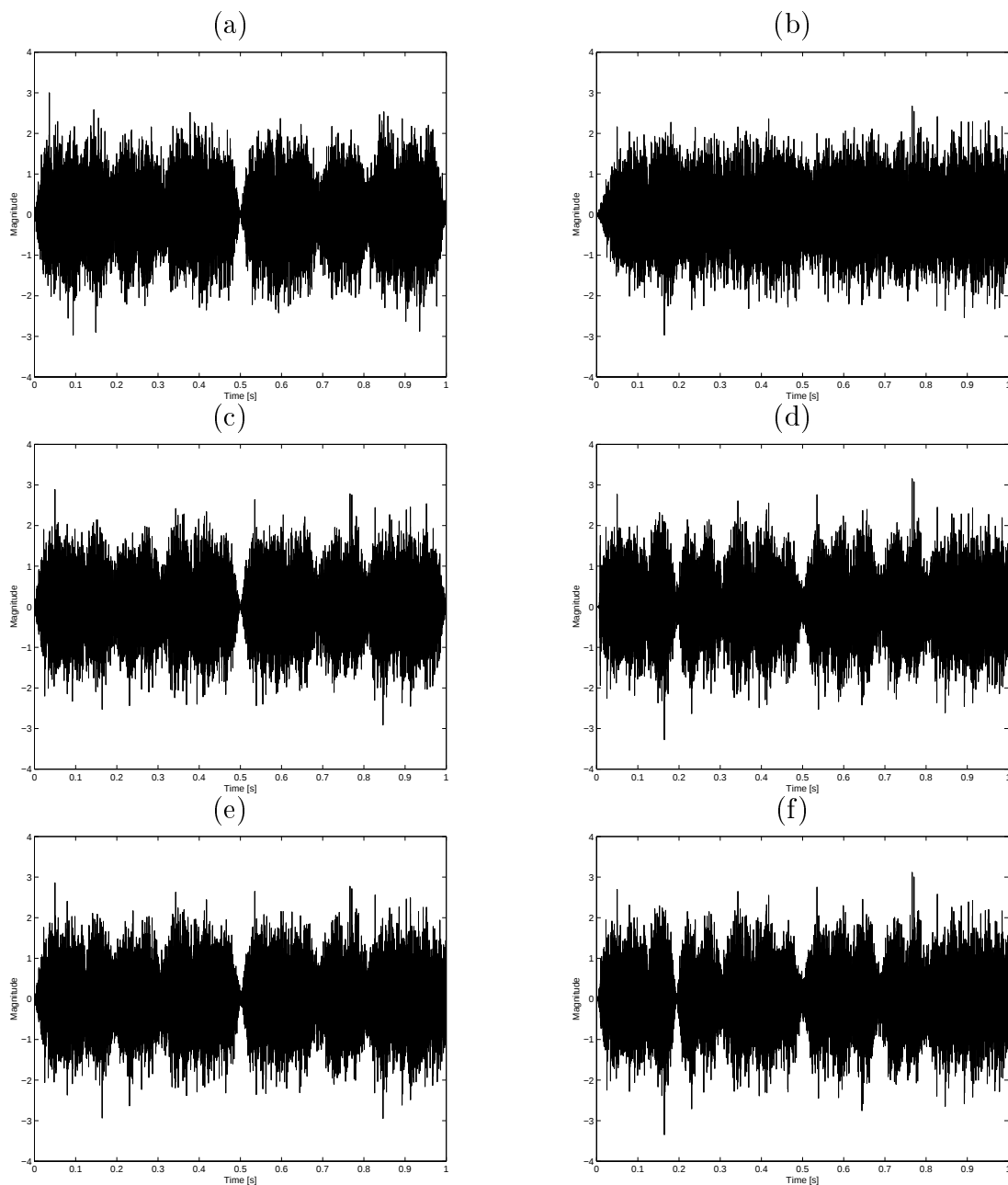


図 5.6: 非共変調信号.(a) 原信号,(b) 残響時間 $T_R = 0.5$ [s] の残響信号,(c) 各帯域における観測信号のパワーエンベロープを既知とした回復信号,(d) 帯域分割による回復信号,(e) 一括処理における観測信号のパワーエンベロープを既知とした回復信号,(f) 一括処理における回復信号

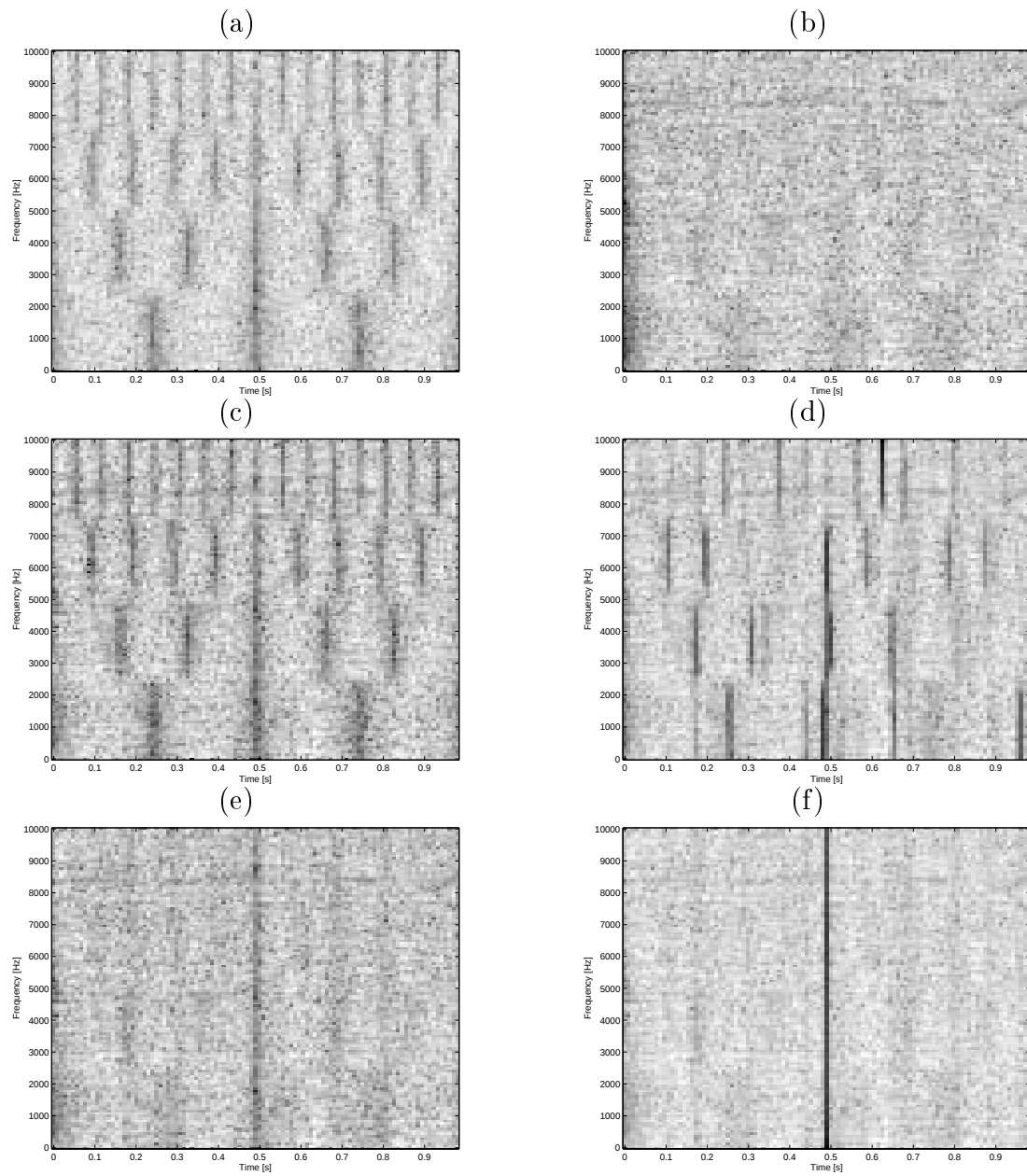


図 5.7: 非共変調信号のスペクトログラム.(a) 原信号,(b) 残響時間 $T_R = 0.5$ [s] の残響信号,(c) 各帯域における観測信号のパワーエンベロープを既知とした回復信号,(d) 帯域分割による回復信号,(e) 一括処理における観測信号のパワーエンベロープを既知とした回復信号,(f) 一括処理における回復信号

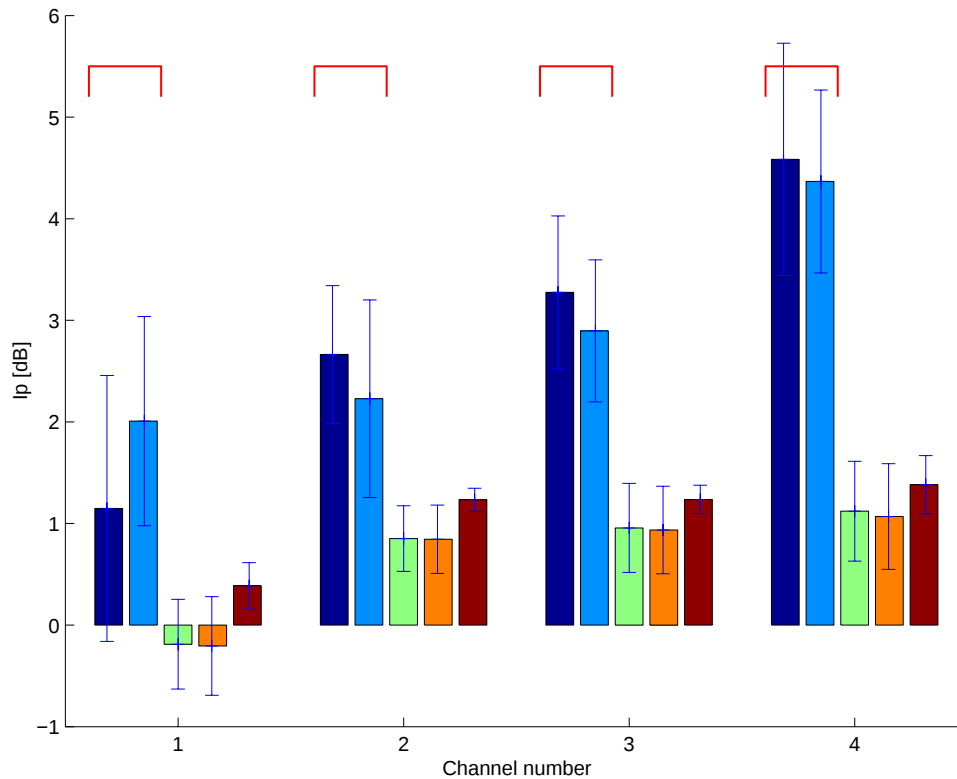


図 5.8: 各チャンネルにおけるパワーエンベロープの回復度.□:分割, なし:一括. 左から (1) 各帯域での残響時間を一定として帯域分割により回復した場合, (2) 各帯域で残響時間を推定して帯域分割により回復した場合, (3) 残響時間を既知として一括処理で回復した場合, (4) 残響時間を推定して一括処理で回復した場合, (5) 一括処理で、観測信号のパワーエンベロープを既知とした場合

次に、定量的な評価を行うため式 (2.24) で示したエンベロープの回復度 I_p [dB] をフィルタバンクの各チャンネルごとに求める。付加した残響の残響時間は 0.5 [s] とし、分析時間 T は約 1.6 [s] である。一括処理においては、パワーエンベロープを回復して回復信号を作成した後、帯域分割をした各々のチャンネルでパワーエンベロープを抽出し評価した。搬送波である白色雑音を 100 種類用い、その平均値と標準偏差を図 5.8 に示す。横軸がフィルタバンクの各チャンネルであり、縦軸がその帯域でのパワーエンベロープ回復度である。棒グラフは左から、(1) 各帯域での残響時間を一定として帯域分割により回復した場合, (2) 各帯域で残響時間を推定して帯域分割により回復した場合, (3) 残響時間を既知として一括処理で回復した場合, (4) 残響時間を推定して一括処理で回復した場合, (5) 一括処理で、観測信号のパワーエンベロープを既知とした場合である。

図 5.8 より帯域分割をしてパワーエンベロープを回復することで、各帯域におけるパワーエンベロープが回復できることがわかる。

5.3.3 考察

図 5.7 より、帯域ごとに包絡線情報が異なる信号を一括処理で回復した場合、全ての帯域で包絡線情報が等しくなるためスペクトルが改善されない。特に、一括処理において観測信号のパワーエンベロープを既知としても包絡線情報が回復されないことから、帯域ごとに包絡線情報が異なる信号に対して一括処理をしても効果がない。これに対して、帯域分割をするとスペクトルが良く改善されることがわかる。図 5.8 より、パワーエンベロープの回復度を定量的に評価したところ、帯域分割をすれば各帯域においてパワーエンベロープが良く回復することがわかる。ここで、 $0 \sim 2.5$ [kHz] の回復度が低い、この帯域における振幅変調周波数 4 [Hz] 以外の成分が多く残存しているため、誤差量が増えパワーエンベロープ回復度が低下したためである。これは、パワーエンベロープを抽出するときにその周波数のみ残すことで改善することができる。

本シミュレーションより、分割したあるチャンネル内において周波数ごとにパワーエンベロープが大きく異なると、その帯域内の包絡線情報が回復されないことは明らかである。そのため、帯域分割をするときに用いるフィルタバンクの帯域幅は十分に注意をする必要がある。

5.4 帯域分割による回復実験

原信号に音声を用いてパワーエンベロープ回復実験を行う。一括処理では各帯域での包絡線情報が正しく回復されないことから、帯域分割して各帯域でのパワーエンベロープを回復する。

5.4.1 分割する帯域幅の決定

音声は各帯域における包絡線情報が異なるため、全帯域の包絡線情報を等しくする一括処理では十分にパワーエンベロープを回復できない。非共変調信号に対して帯域分割を行い回復処理をすることで、各帯域でのパワーエンベロープが良く回復されることから、音声においても帯域分割をして各々の帯域でパワーエンベロープを回復する必要があると考えられる。しかし、帯域分割を行うことで搬送波の周波数成分が制限されるため、原信号と室内インパルス応答の搬送波が無相関ではなくなり式 (2.17) が成立しなくなる。このように、音声を帯域分割してパワーエンベロープを回復する場合、その帯域幅をどのように決定するかが重要である。

ここで、

1. 音声は各帯域における包絡線情報が異なる
2. パワーエンベロープ逆フィルタ処理は共変調として回復する

ことから、分割する帯域幅は音声における各帯域のパワーエンベロープが共変調している

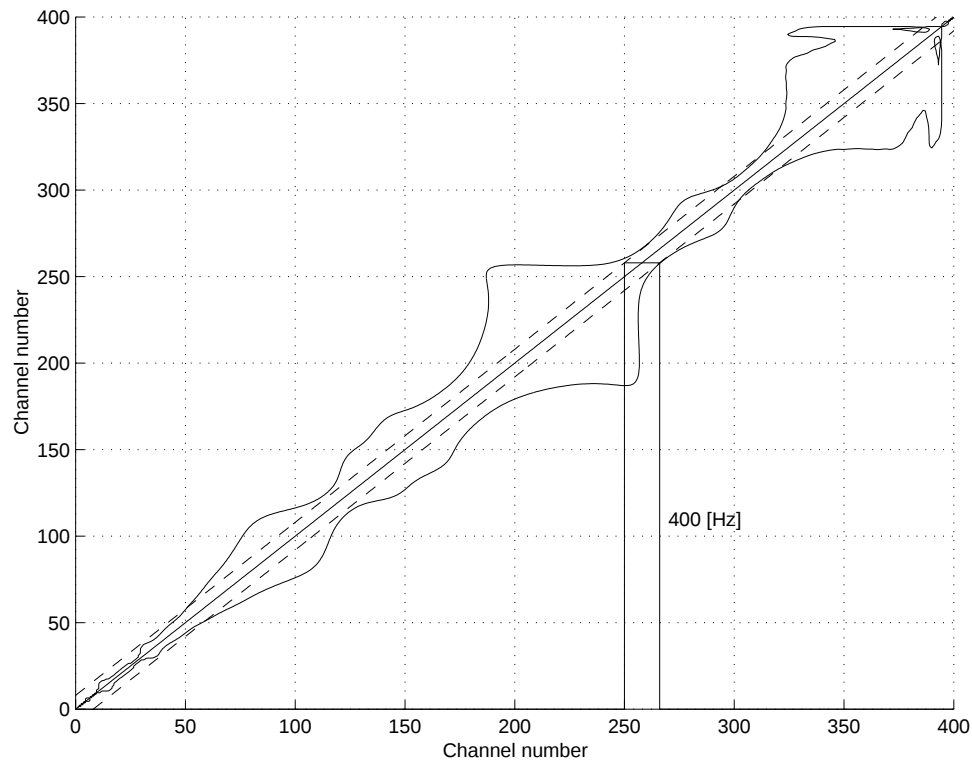


図 5.9: 各帯域間でのパワーエンベロープの相関値 (0.95 以上).

と考えられる帯域幅を調べるにより決定できると考えられる。このようにすることで、帯域分割をすることにより MTF の理論が破綻する影響を抑えることができる。

各帯域のパワーエンベロープが共変調している周波数領域を調査するため、パワーエンベロープの周波数領域を考慮して、残響音声を帯域幅 25 [Hz] 等間隔のフィルタバンクで分割して各帯域ごとにおけるパワーエンベロープの相関値を求めた。音声は ATR 音声データベース mau “相変わらず” を用いた。但し、残響信号の残響時間は 0.5 [s] である。

図 5.9 の横軸は各チャンネルを表し、縦軸のチャンネルとの相関値が 0.95 以上の部分を示している。右上がりの実直線は相関値が 1 である。

相関値が 0.95 以上の帯域では共変調していると考える。図 5.9 において、実線で囲まれた領域内で相関値が 1 の地点を中心に横軸と平行な距離が、共変調していると見なせる帯域幅である。これより、各帯域において共変調と見なせる帯域幅は異なることがわかる。本研究ではフィルタバンクを簡単に構成するため定帯域のフィルタバンクを当てはめる。そこで、相関値が 1 を表す実直線に平行で共変調と見なせる領域を多く含むように点線を描いた。その結果、帯域幅が約 400 [Hz]、サンプリング周波数が 20 [kHz] であるため分割数が 25 個の定帯域フィルタバンクが最適と考えられることがわかる。以後、帯域分割には帯域幅 400 [Hz] の定帯域フィルタバンクを用いる。

5.4.2 検証シミュレーション

音声データにはATR 音声データベース_{mau}の“相変わらず”を用いた。帯域幅 400 [Hz] の定帯域フィルタバンクにより残響音声を帯域分割して、各々の帯域においてパワーエンベロープを抽出する。抽出したパワーエンベロープに対してパワーエンベロープ逆フィルタ処理を施し、パワーエンベロープを回復する。パワーエンベロープの回復度を定量的に評価するため、式 (2.24) で示したパワーエンベロープ回復指標 I_p [dB] を各帯域ごとに求める。

5.4.3 検証結果

残響時間 0.5 [s] の残響を付加した残響音声に対して、帯域幅 400 [Hz] の等帯域フィルタバンクにより帯域分割した。分割した各帯域に対してパワーエンベロープ逆フィルタ処理を施し、パワーエンベロープの回復をした。パワーエンベロープの回復では、全帯域で残響時間が一定として回復した場合と各帯域ごとに残響時間を推定して回復した場合を示す。回復音声のスペクトログラムを図 5.10 に示す。(a) は原音声、(b) は残響時間 0.5 [s] の残響を付加した残響音声、(c) は全ての帯域で残響時間が同じとしたときの回復音声、(d) は残響時間を推定したときの回復音声、(d) は観測信号のパワーエンベロープを既知としたときの回復音声である。

一括処理により回復した図 5.1 と比較すると、各帯域において包絡線情報を良く回復できていることがわかる。全帯域の残響時間が 0.5 [s] で一定であるとしたときの (c) では、回復音声における各帯域のパワーエンベロープがばらついており、ひずみが見られる。各帯域で残響時間を推定した (d) では、(c) のようなひずみは見られない。ここで、各帯域におけるパワーエンベロープ回復指標を図 5.11 に示す。

(a) は全帯域における残響時間が一定であると仮定して回復した場合、(b) は各帯域において残響時間を推定して回復した場合である。これより、全帯域で残響時間が一定であるとしたときパワーエンベロープ回復指標は、高域を除いて向上していないことがわかる。残響時間を推定した場合、全ての帯域でパワーエンベロープの改善が見られ全帯域で残響時間を一定とした場合と比べ、パワーエンベロープ回復指標が向上している。

5.4.4 考察

図 5.11 より、全帯域で残響時間が一定であるとして回復するよりも、帯域ごとに残響時間を推定した方がパワーエンベロープをより回復する。これは、回復のために用いる最適な残響時間が、帯域ごとに異なるためと考えられる。また、実環境下において観測される残響音声は、全帯域で残響時間が同じと考えるより帯域ごとに異なると考えた方が自然である。そのため、帯域ごとに残響時間の推定をすることが有効に働くと考えられる。また、残響時間の推定法はパワーエンベロープの相関値と SNR が比較的高い残響時間を

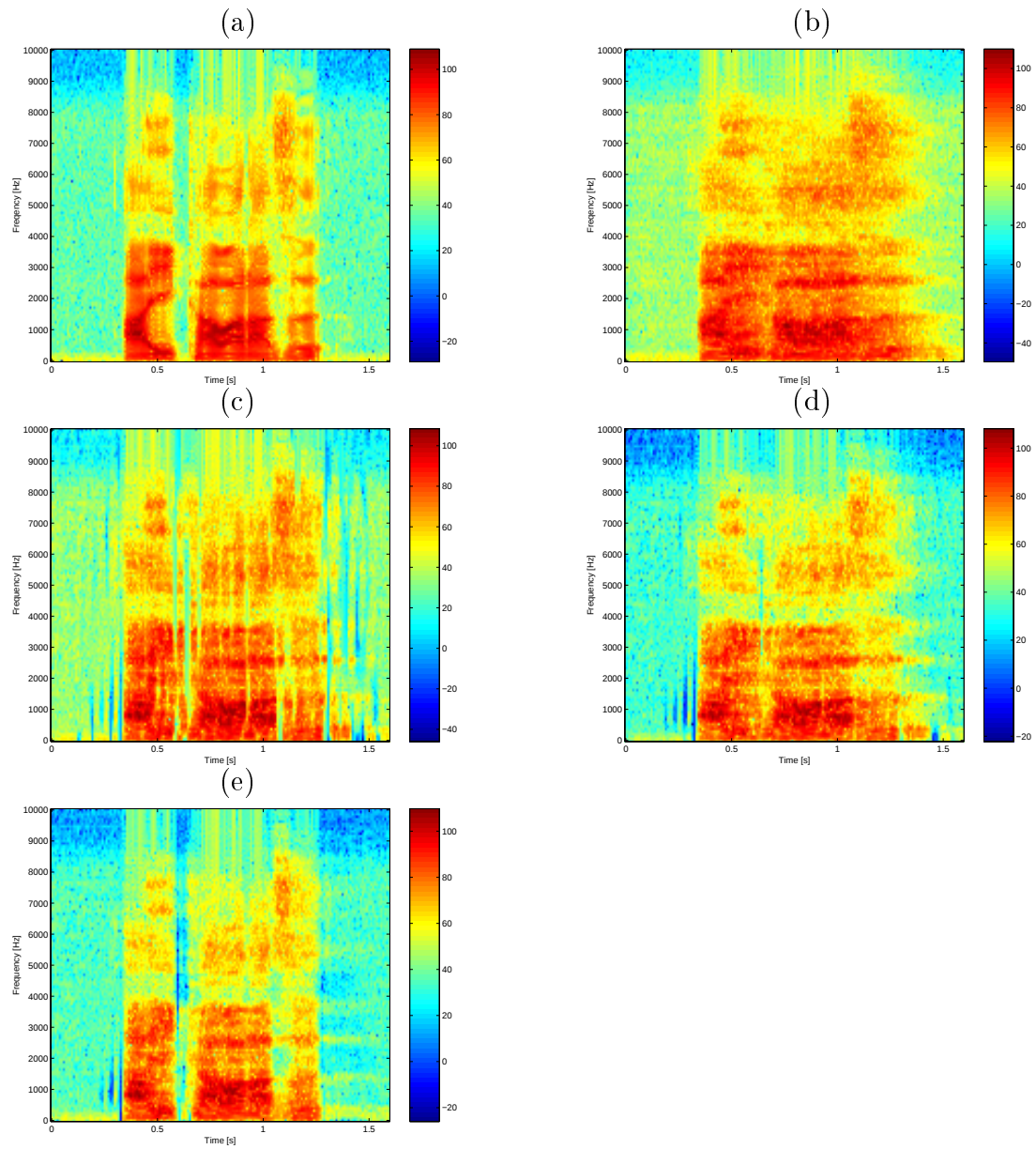


図 5.10: 帯域分割により回復した音声のスペクトログラム (a) 原音声, (b) 残響時間 0.5 [s] の残響を付加した残響音声, (c) 全ての帯域で残響時間が同じとしたときの回復音声, (d) 残響時間を推定したときの回復音声, (e) 観測信号のパワーエンベロープを既知としたときの回復音声.

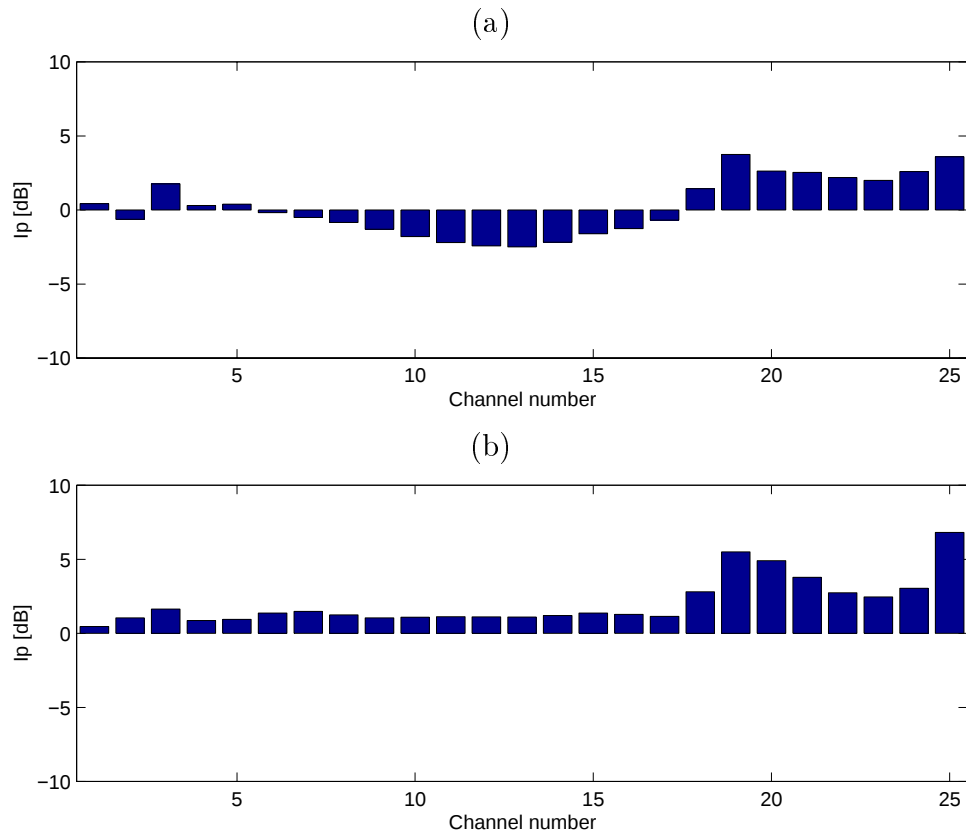


図 5.11: 各チャネルにおけるパワーエンベロープ回復指標.(a) 残響時間を一定とした場合,(b) 残響時間を推定した場合

推定したため、パワーエンベロープ回復指標が向上したと考えられる。これより、帯域分割をすることで音声のパワーエンベロープを回復することができ、実環境下で観測された残響音声に対しても帯域ごとに残響時間を推定すれば、パワーエンベロープを回復することができると考えられる。

図 5.11 より、低域のパワーエンベロープ回復指標が高域に比べ低いことがわかる。これは、図 5.9 からわかるように低域のパワーエンベロープが帯域ごとに激しく変動しており、共変調しているとした帯域幅が不適切であったためである。この問題は、各帯域ごとに適切な帯域幅をもつフィルタバンクを構成することで解決できると考えられる。

図 5.11 に示したパワーエンベロープ回復指標は各帯域でパワーエンベロープが回復している。しかし、室内インパルス応答の搬送波である白色雑音の種類を変化させると、パワーエンベロープの回復がほとんどされない場合がある。その例を図 5.12 に示す。

図 5.12 より、帯域ごとに残響時間を推定してもパワーエンベロープが回復されていないことがわかる。これは、音声の搬送波成分と室内インパルス応答の搬送波の関係によるものである。

MTF の理論が成り立つ場合、残響音声のパワーエンベロープは原音声のパワーエンベ

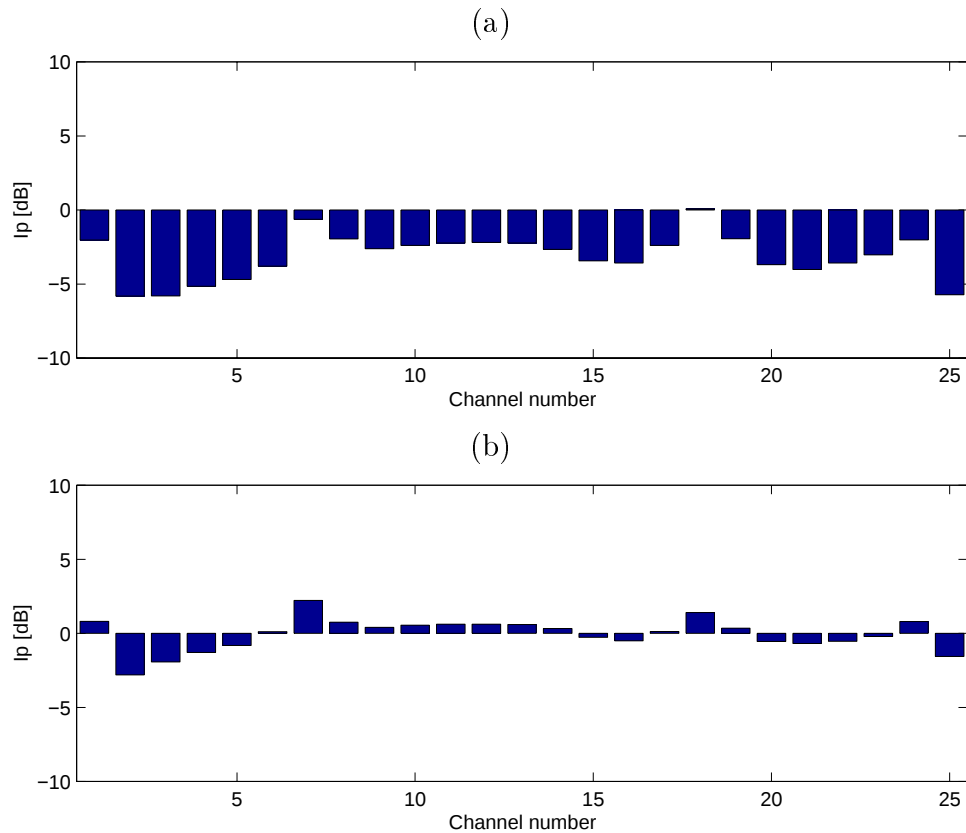


図 5.12: 各チャネルにおけるパワーエンベロープ回復指標 (回復されない場合). (a) 残響時間を一定とした場合, (b) 残響時間を推定した場合

ロープに比べ時間的に遅れ、ピークが低くならない(図 5.13(a))。一括処理でのパワーエンベロープ抽出において、全帯域で抽出処理をしているため理論上 MTF の理論が成立するはずである。しかし室内インパルス応答の搬送波である白色雑音の種類によりピークが低くならないことがある(図 5.13(b))。これは、音声を搬送波と振幅エンベロープの掛け合わせと考えたとき、音声の搬送波と室内インパルス応答の搬送波成分である白色雑音とが無相関ではなくなったため、MTF の理論が破綻した影響によりピークが低くならなかったと考えられる。全帯域を用いてパワーエンベロープを抽出した場合においてピークが低くないため、この影響が帯域分割した場合にも現れている。1.6 ~ 2 [kHz] でのパワーエンベロープを図 5.14 に示す。

この問題は、MTF の理論が破綻した影響により残響音声のパワーエンベロープが原音声のパワーエンベロープと室内インパルス応答のパワーエンベロープとの畳込みで定義できるなることが原因である。MTF の理論の破綻が残響音声パワーエンベロープに与える影響を調査することで、その影響を補正してパワーエンベロープの回復が可能であると考えられる。

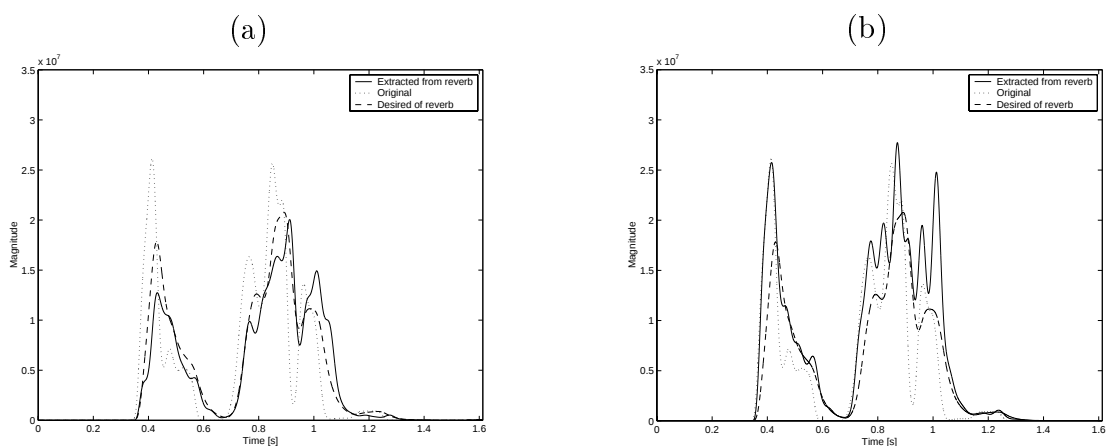


図 5.13: 異なる搬送波白色雑音における抽出したパワーエンベロープ (全帯域). (a) MTF の理論が成立しているとき, (b) MTF の理論が成立していないとき (実線: 残響音声のパワーエンベロープ, 点線: 原音声のパワーエンベロープ, 破線: 残響音声パワーエンベロープの理論値)

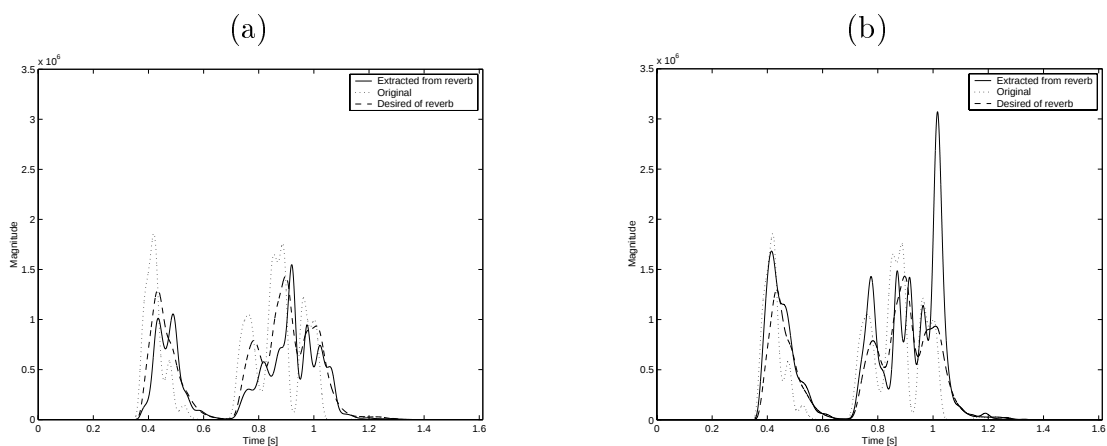


図 5.14: 異なる搬送波白色雑音における抽出したパワーエンベロープ (1.6 ~ 2 [kHz]). (a) MTF の理論が成立しているとき, (b) MTF の理論が成立していないとき (実線: 残響音声のパワーエンベロープ, 点線: 原音声のパワーエンベロープ, 破線: 残響音声パワーエンベロープの理論値)

第6章 まとめ

6.1 本研究で明らかにされたことの要約

本研究では、残響波形の情報のみから残響時間を推定して残響波形パワーエンベロープの回復をした。残響波形のパワーエンベロープを回復するパワーエンベロープ逆フィルタ処理の有効性を確認し、明らかになった問題点を解決した。

パワーエンベロープの抽出法では、ヒルベルト変換, パワー, 包絡線検波, 集合平均, 集合平均+ヒルベルト変換, 集合平均+パワー, 集合平均+包絡線検波の7つの手法を提案してそれぞれを比較した。その結果、ヒルベルト変換による方法が最も抽出精度が高いことがわかった。

残響時間の推定では、負の面積による方法, 変調伝達関数による方法, 高周波数成分による方法を提案して、それぞれを比較検討した。その結果、負の面積による方法は残響信号パワーエンベロープの情報のみから残響時間を推定できることがわかった。また、推定した残響時間を用いてパワーエンベロープの回復をすれば、回復度が向上することもわかった。

音声のパワーエンベロープ回復では、搬送波が調波複合音のときについて検証を行い、提案したパワーエンベロープ抽出法と残響時間の推定法が有効であることを確認した。また、MTFの理論が成立するとき分割した帯域ごとに残響時間を推定してパワーエンベロープを回復することで、各帯域における音声の包絡線情報が良好に回復できることを示した。

最後に図6.1に本研究全体のブロック図を示す。太枠の処理が本研究で検証した部分である。

6.2 今後の課題

残響時間推定の誤差

付加された残響の残響時間が長くなると、残響時間の推定誤差が増加する。これは、残響波形から抽出したパワーエンベロープの周波数成分による影響である。パワーエンベロープを抽出するとき、理想低域フィルタに近い低域フィルタで必要な周波数成分だけを取り出すことで、残響時間の推定誤差は軽減できる可能性がある。

搬送波成分の抽出

回復信号を作成するとき、残響信号をそのパワーエンベロープで除算することにより搬送波成分を作成している。このため、回復信号には残響信号の影響が残ると考えられる。また、パワーエンベロープの値が0の場合、除算できない。これより、搬送波成分を抽出する手法の構築が必要である。

帯域分割処理

帯域分割することにより音声のパワーエンベロープを回復する場合、音声の搬送波成分と室内インパルス応答の搬送波成分が無相関であれば、残響波形のパワーエンベロープが回復できる。しかし、実環境下においてこのような条件を満たすとは限らない。また、搬送波成分の無相関性は分割する帯域幅にも関係してくる。このため、MTFの理論と分割する帯域幅との関係について十分に検討する必要がある。

フィルタバンクの構成

本研究では、定帯域フィルタバンクを用いた。しかし、帯域ごとに共変調と考えられる周波数帯域幅が異なるため、定帯域フィルタバンクでは適切な分割とは言えない。また、共変調と考えられる周波数帯域幅は音声の種類により異なる。そこでフィルタバンクを構成するとき、音声を観測することにより共変調していると考えられる部分を自動的に求め、最適な帯域幅を持つようにすることでより音声の包絡線情報が回復すると考えられる。

時間方向の分割

本研究では周波数方向の分割である帯域分割を行い、残響波形のパワーエンベロープを回復した。実環境に適用する場合、逐次処理をすればと考えれば時分割によるパワーエンベロープの回復に関して調査する必要がある。

時間波形の回復

本研究は、振幅の特徴を反映するMTFの理論に基づいているため位相情報を無視している。そのため、時間波形の復元はできない。波形レベルでの復元をする場合には、残響環境における位相情報について理論的な検討が必要である。

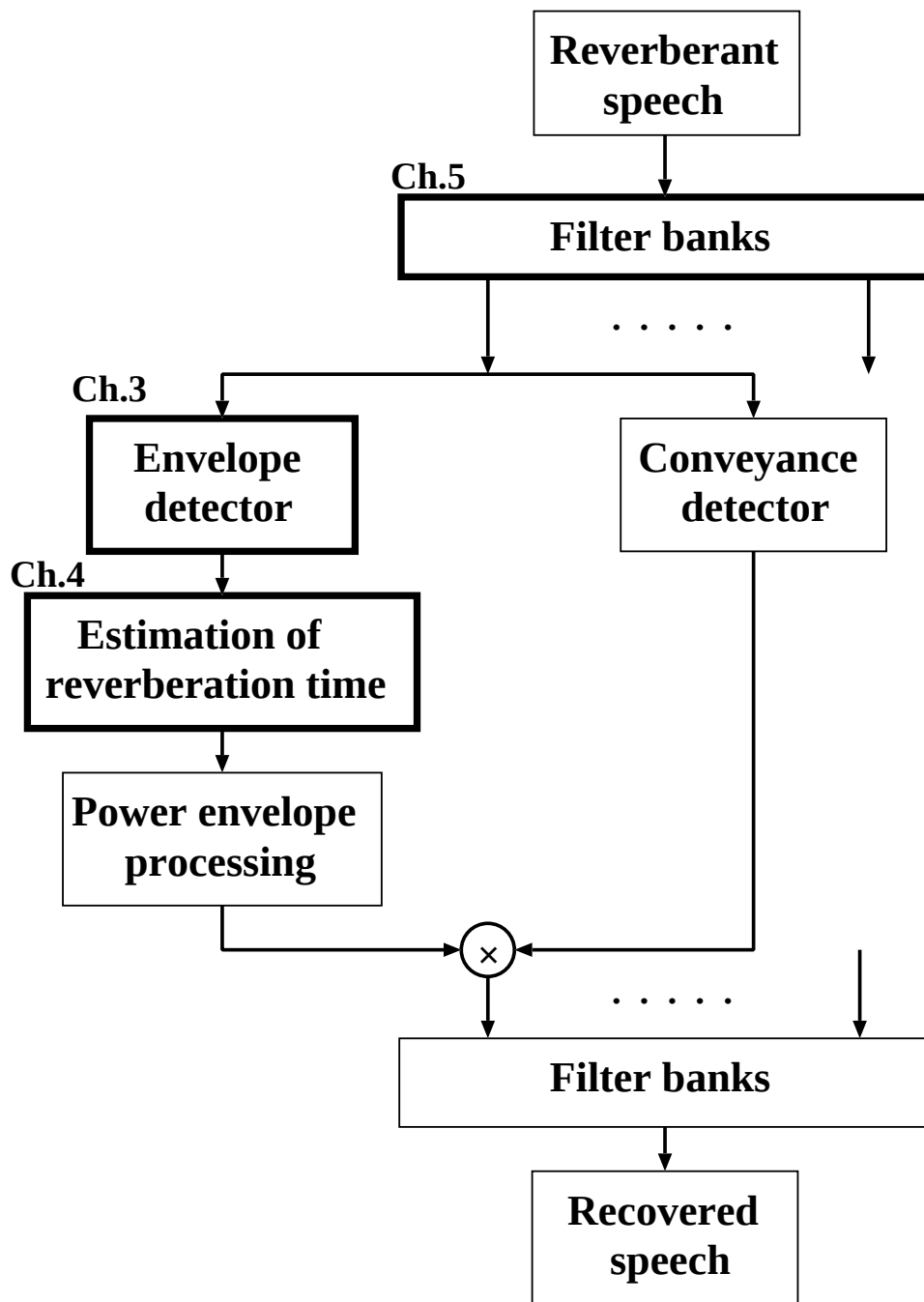


図 6.1: ブロックダイアグラム.

謝辞

本研究を進めるにあたり、多大な御助言を頂き熱心な御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。本研究を進めるにあたり、多大な御助言を頂き熱心な御指導をして頂きました党建武助教授に心から感謝致します。本研究に関して、多大な御助言、御討論をして頂いた鵜木祐史助手に心より感謝致します。本研究に関して、多大な御助言をして頂いた博士後期課程の伊藤一仁氏、石本祐一氏、西本博則氏に心より感謝致します。また、本研究を進めるにあたり有意義な討論並びに、有益な助言を賜った赤木・党研究室の皆様方に心より感謝致します。

関連図書

- [1] M. Miyoshi and Y. Kaneda, “Inverse filtering of room acoustics”, IEEE Trans. ASSP, Vol.36, No.2, pp.145-152, Feb. 1988.
- [2] H. Wang and F. Itakura, “Realization of acoustic inverse filtering through multi-microphone sub-band processing”, IEICE Trans. Fundamentals, Vol.E75-A, No.11, pp.1474-1483, Nov. 1992.
- [3] 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 A, Vol. J81-A, No.10, pp.1323-1330, 1998.
- [4] 広林茂樹, 山淵龍夫, “帯域分割を用いたパワーエンベロープ逆フィルタ処理の残響抑圧効果”, 電子情報通信学会論文誌 A, Vol. J83-A, No.8, pp.1029-1033, 2000.
- [5] 広林茂樹, 小池恒彦, 東山三樹夫, “最小位相成分とパワーエンベロープに着目した残響音場による音源波形回復”, 電子情報通信学会技術研究報告, EA93-81, pp.29-36, Dec. 1993.
- [6] M.R.Schroeder, “Modulation Transfer Functions: Definition and Measurement”, Acustica, Vol. 49, pp.179-182, 1981.
- [7] T.Houtgast, H.J.M. Steeneken, and R.Plomp, “Predicting speech intelligibility in room acoustics”, Acustica, Vol.46, pp.60-72, 1980.
- [8] T.Houtgast and H.J.M. Steeneken, “A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria”, J. Acoust. Soc. Am. Vol. 77, No.3, March. 1985.
- [9] Stephen T. Neely and Jont B. Allen, “Invertibility of a room impulse response”, J. Acoust. Soc. Am. Vol. 66, No. 1, July 1979.
- [10] N. Kanedera, et. al., “On the importance of various modulation frequency for speech recognition”, Proc. Eurospeech, pp.1079-1082, Rhodes, Greece, Sept. 1997.

- [11] 金寺, 荒井, 船田, “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 D-II, Vol. J84-D-II, No.7 pp.1261-1269, 2001.
- [12] T.Arai, M.Pavel, H.Hermansky, and C.Avendano, “Intelligibility of Speech with Filtered Time Trajectories of Spectral Envelopes”, In Proc. of the ICSLP, pp.2490-2493, Philadelphia, October, 1996.
- [13] Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1988). “Numerical Recipes in C,” Cambridge University Press, Cambridge.
- [14] 三浦, 赤木, “振幅変調マスキングの知見に基づいた音源分離に関する研究”, 北陸先端大, 修士論文, 1999.
- [15] 日本音響学会 “音・振動のスペクトル解析”, 金井 浩著, コロナ社, pp.293-306, 1999.
- [16] CQ 出版社, “デジタルフィルタ設計入門”, 尾知 博著, pp.113-158.