JAIST Repository
https://dspace.jaist.ac.jp/

Title	変調伝達関数に基づく音声信号処理(1)パワーエンベロープ逆フィルタ処理の原理とその応用について
Author(s)	鵜木，祐史
Citation	信号処理，12(5)：339-348
Issue Date	2008-09
Type	Journal Article
Text version	author
URL	http://hdl.handle.net/10119/7754
Rights	Copyright (C) 2008 信号処理学会．鵜木祐史，信号処理，12(5)，2008，339-348．
Description	

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

変調伝達関数に基づく音声信号処理 (1)
—パワーエンベロープ逆フィルタ処理の原理とその応用について—
Speech Signal Processing Based on the Concept of Modulation Transfer Function (1)
— Basis of Power Envelope Inverse Filtering and Its Applications —

鵜木祐史
Masashi Unoki

1. はじめに

私達は，日常，何不自由なく音声を介してコミュニケーションをとっている。しかし，読者はこんな経験をしたことはないだろうか。例えば，お風呂場や教会など音が非常に響く環境（残響環境）や，人で賑わっている雑多な場所，交通量の多い場所といった非常に騒がしい環境（騒音環境）では，静寂な環境に比べて非常に音を聴き取り難く，いつもと同じように簡単に会話をできないと感じたことである。これは，身の回りの音場環境の影響により，音声が歪んだため，音声知覚に重要な情報が欠落したことによるものである[1]。

このような音声コミュニケーションの難しさを評価する尺度として，音声明瞭度，単語・文章了解度が利用されている [2]。前者は無意味音節を発声したとき受聴者がその何%を正しく聞き取れたかを，後者は沢山の有意味単語を発声したとき受聴者が正しく聞き取れた単語数の割合を示すものである。これらの尺度は，音声情報伝達を議論するときに，よく利用されるものであるが，同時に室内音響学と関係して議論されるとき，音声レベル，騒音レベル，残響時間等の物理量との関連を見出そうとする検討も古くから行われている [3]。

代表的なものとして，Houtgast と Steeneken によって提唱された変調伝達関数（Modulation Transfer Function: MTF）に基づく音声明瞭度予測理論 [4, 5, 6] がある。これは，音場内において，音声波形の時間的な包絡線情報（以後，エンベロープと呼ぶ）が残響や雑音によって変形することに着目し，100%振幅変調した正弦波を利用して MTF の減衰量から音声伝達指標（Speech Transmission Index: STI）を予測するものである。この方法は，その後，簡易測定法である RASTI[7] として提案され，現在でも標準的な方法として利用されている。STI/RASTI の方法は，理論的に明解であり，実用上多くの利点をもつため，講演会場など室内音響設計 [2, 8] にも役立っている。しかしながら，この方法は決して万能であるわけではなく，(1) 音場の時間構造・空間構造を反映していないことや (2) 音源（音声）の物理特性を反映していないことから，音声明瞭度予測に対して適用限界があることが示唆されている [2]。

Houtgast と Steeneken が提唱した音声明瞭度予測理論は，室内音響を拡散音場と仮定しているため [8]，上記のように，その適用限界があることは間違いない。しかしながら，室内音響伝達系を入出力の強度情報の関係と残響・雑音に対する MTF を明解に関係づけた点は，大きな業績であり，評価されることであろう。また，この考えは，他の音声信号処理で残されている諸問題を解決するために利用することもできる。例えば，室内の残響の影響を受けた音声を伝達系を測定せずに回復する方法 [9, 10, 11] や，残響環境下での音声の基本周波数推定方法 [12] がある。最近では，室内の残響時間をブラインド推定する方法 [13] や異なる二つの音場空間を考慮した音場再生法 [14] も提案されている。

本論文は，合計 3 回のシリーズで構成される。これらでは，著者が関係した研究分野（残響環境下の音声信号処理）を中心に，MTF を利用した音声信号処理を解説する。本稿では，まず，Houtgast と Steeneken が示した MTF の概念を解説するとともに，その概念に基づいたパワーエンベロープ逆フィルタ法を紹介する。

2. 変調伝達関数（MTF）

変調伝達関数（MTF）の概念は，音声明瞭度を評価するための室内音響の測定として，Houtgast と Steeneken[4, 5, 6] によって導入された。彼らの着眼点は，室内音響において，残響や単一エコー，妨害雑音の変化によって得られる明瞭度のスコアと MTF から

[1] 聴覚の優れた能力の一つに音韻修復現象 [1] がある。これは雑音等で欠落した情報から情報を知覚的に補正・修復することで，正しく音声を知覚できるという現象であるが，言語や先験情報に依存しており，高次からの処理が重要と考えられている。同様に雑音・残響環境下においてもこの能力が発揮されるが，この現象が説明されるような状況とは異なり，音声全体にわたって歪みが存在するため，誤った修復を行い，結果として異聴を引き起こすことが多いと思われる。

北陸先端科学技術大学院大学 情報科学研究科
923-1292 石川県能美市旭台 1-1
School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
E-mail: unoki@jaist.ac.jp

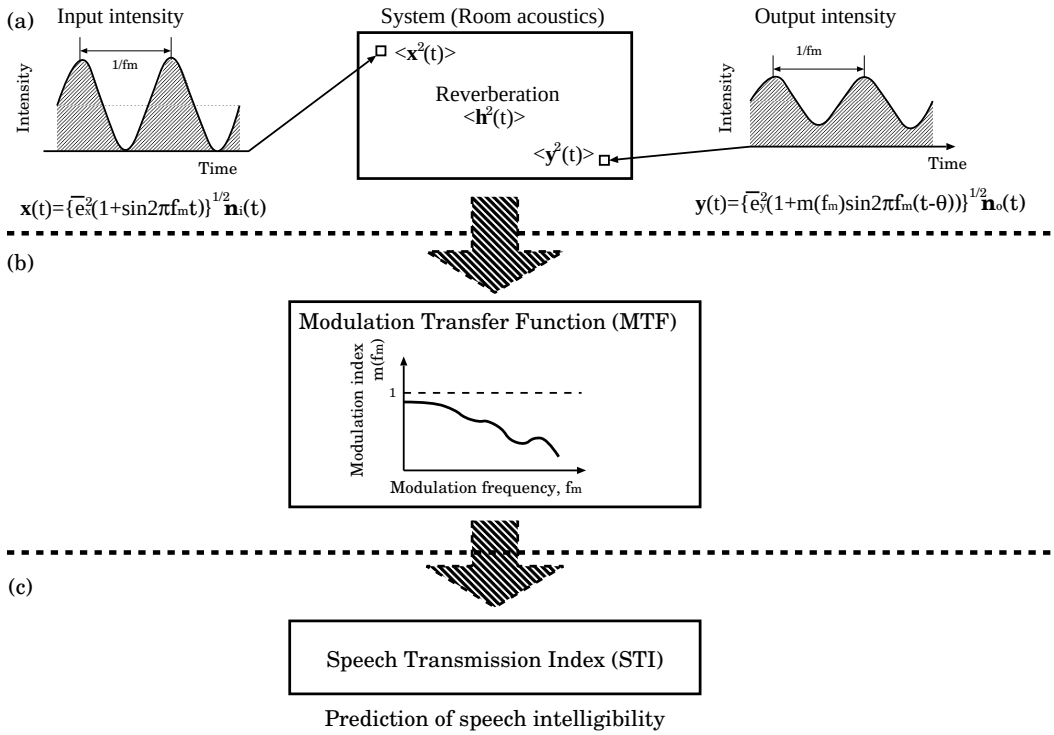

図 1　Houtgast と Steeneken によって提案された音声明瞭度予測理論の概要図。(a) 室内音響における入出力信号の強度情報とその伝達関数。(b) 導出された変調伝達関数（MTF）$m(f_m)$。(c) MTF から導出される音声伝達指標（STI）と音声明瞭度の予測

Fig. 1 Schematic illustration of prediction of speech intelligibility proposed by Houtgast and Steeneken: (a) Relationship between intensity of input and output signals and its transfer function, (b) modulation trasnfer funcion (MTF), and (c) speech transmission index (STI) derived from MTF

導出される指標が非常に高い相関関係にあることにある。彼らが想定している室内音響は拡散音場であるが，次のような二つの場合に分類している。(a) 残響に関係する伝達関数は，直接経路の他に，壁や床など様々な反射経路（時間差あり）の総和で構成されること，(b) 妨害雑音に関係する伝達関数は，ある点（音源）から室内のある一点までの経路上における妨害量を表すものである。そのため，彼らは，室内を伝達系と見たときの入力，出力の強度変化に着目し，その中の変調度の変化を MTF と定義して，音声明瞭度に関係する指標 STI を予測する体系を作った。本稿では，MTF に特化するため，STI の導出法については触れないが，詳細については，[4, 5, 6] を参照されたい。

まず，彼らは，室内音響における入出力の強度変化を

$$\textbf{Input:} \quad \overline{I_i}(1+\cos(2\pi f_m t)) \tag{1}$$

$$\textbf{Output:} \quad \overline{I_o}(1+m(f_m)\cos(2\pi f_m(t-\theta))) \tag{2}$$

と定義した。ここで，$\overline{I_i}$ と $\overline{I_o}$ はそれぞれ入出力の強度，f_m は変調周波数，θ は位相を表す。入力の強度をパワーエンベロープとみなしたときに，ちょうど変調度 $m(f_m)=1$（100%振幅変調）の余弦波をもつものとする。出力は同様の形態を取るが，余弦波の振幅が雑音や残響の影響を受け，ちょうど $m(f_m)$ 倍に変化し，これが振幅変調で見たときの変調度に相当することから，MTF と呼ばれる。

次に $m(f_m)$ の導出を説明するために，入力信号 $x(t)$ と出力信号 $y(t)$ のパワーエンベロープを

$$e_x^2(t) = \overline{e_x^2}(1+\cos(2\pi f_m t)) \tag{3}$$

$$e_y^2(t) = \overline{e_y^2}(1+m(f_m)\cos(2\pi f_m t)) \tag{4}$$

と定義し（便宜上，$\theta=0$），次式のように一般化する。

$$\mathbf{x}(t) = e_x(t)\mathbf{n}_x(t) \tag{5}$$

$$\mathbf{y}(t) = e_y(t)\mathbf{n}_y(t) \tag{6}$$

但し，$\mathbf{n}_x(t)$ ならびに $\mathbf{n}_y(t)$ は，白色ガウス雑音の特性を有するランダム変数である。このランダム雑音には次式の特性がある（x と y の添字は省略する）。

$$<\mathbf{n}(t)\mathbf{n}(\tau)> = \delta(t-\tau) \tag{7}$$

ここで，$<\cdot>$ は集合平均を表す [15]。この特性を活用し，$\mathbf{x}(t)$ と $\mathbf{y}(t)$ の 2 乗集合平均を求めると，

$$<\mathbf{x}^2(t)> = e_x^2(t) \tag{8}$$

$$<\mathbf{y}^2(t)> = e_y^2(t) \tag{9}$$

を得る（導出過程については付録 1 を参照）。

2.1 残響環境 $h(t)$ における MTF

残響環境における出力信号 $\mathbf{y}(t)$ は入力信号 $\mathbf{x}(t)$ と室内インパルス応答 $h(t)$ の畳み込み演算から得られる。

$$\mathbf{y}(t) = \int_0^\infty h(\tau)\mathbf{x}(t-\tau)d\tau \tag{10}$$

両辺を 2 乗した集合平均 $<\cdot>$ を求めると，次式を得る（詳細な導出過程については付録 2 を参照）。

$$<\mathbf{y}^2(t)> = e_y^2(t) = \frac{\overline{e_x^2}}{\alpha}\left[1+\frac{\beta}{\alpha}\cos(2\pi f_m t)\right] \tag{11}$$

但し，$\alpha = \int_0^\infty h^2(\tau)d\tau$，$\beta = \int_0^\infty h^2(\tau)\exp(-j\omega_m\tau)d\tau$ である。ここで，式 (4) との対応を見れば，$m(f_m) = \beta/\alpha$ がわかり，次式に示す複素表現の MTF を得る。

$$\mathbf{m}(\omega) = \frac{\int_0^\infty h^2(t)\exp(-j\omega t)dt}{\int_0^\infty h^2(t)dt} \tag{12}$$

これは $h^2(t)$ に対する一種の Fourier 変換を表している。

次に，式 (12) のインパルス応答 $h(t)$ を次式のように定義する。これは，室内音響特性の統計的近似として良く知られた Schroeder のインパルス応答 [16] である。

$$\mathbf{h}(t) = a\exp\left(-\frac{6.9t}{T_R}\right)\mathbf{n}_h(t) = e_h(t)\mathbf{n}_h(t), \quad t \geq 0 \tag{13}$$

但し，a は振幅項，T_R は T_{60}（$h(t)$ のパワーがそのピークからちょうど 60 dB 減衰するとるまでに要する時間）で定義される残響時間である [8, 16, 17]。このとき，MTF は，式 (12) に式 (13) を代入することで得られる。

$$m(f_m) = |\mathbf{m}(2\pi f_m)| = \left[1+\left(2\pi f_m\frac{T_R}{13.8}\right)^2\right]^{-1/2} \tag{14}$$

図 2 は，変調周波数 f_m（時間的なパワーエンベロープにおける主要な周波数）の関数とした式 (14) の MTF $m(f_m)$ の特性図を示す。これらの曲線は，$T_R = 0.1, 0.3, 0.5, 1.0, 2.0$ s と $\omega = 2\pi f_m$ としたときのものである。$m(f_m)$ は，100%振幅変調したときの f_m に対する変調度と解釈することもできる。これらの曲線は，T_R と f_m に依存しており，その概形からわかるように，残響環境の MTF は一種の低域通過フィルタ特性を示している。

2.2 雑音環境 $w(t)$ における MTF

雑音環境として，定常な加法性雑音 $\mathbf{w}(t)$ を考える。

$$\mathbf{w}(t) = e_w(t)\mathbf{n}_w(t) \tag{15}$$

ここで，図 1 における入出力関係は

$$\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{w}(t) \tag{16}$$

である。前節にならい，両辺の 2 乗集合平均を取ると

$$\begin{aligned} <\mathbf{y}^2(t)> &= <\mathbf{x}^2(t)> + <\mathbf{w}^2(t)> \\ e_y^2(t) &= e_x^2(t) + e_w^2(t) \\ &= \left(\overline{e_x^2}+\overline{e_w^2}\right)\left(1+\frac{\overline{e_x^2}}{\overline{e_x^2}+\overline{e_w^2}}\cos(2\pi f_m t)\right) \end{aligned}$$

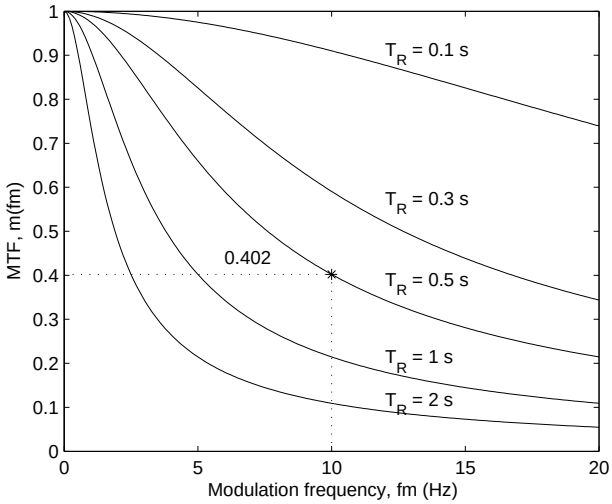

図 2　様々な残響時間（$T_R = 0.1, 0.3, 0.5, 1.0, 2.0$ s）における変調伝達関数 $m(f_m)$ の特性

Fig. 2 Theoretical curves representing the MTF, $m(f_m)$, for various conditions with $T_R = 0.1$, 0.3, 0.5, 1.0, and 2.0 s

を得る。但し，$e_w^2(t)$ は定常と仮定され，$e_w^2(t) = \overline{e_w^2} = \frac{1}{T}\int_0^T e_w^2(\tau)d\tau$ とする。ここで T は信号長である。このとき，式 (3) と式 (4) の関係を眺めると，余弦波にかかる $m(f_m)$ は f_m に依存せず，次式のように一定な減衰を表すことがわかる。

$$m(f_m) = \frac{\overline{e_x^2}}{\overline{e_x^2}+\overline{e_w^2}} = \frac{1}{1+10^{-(S/N)/10}} \tag{17}$$

但し，S/N は $10\log_{10}\overline{e_x^2}/\overline{e_w^2}$ (dB) である。例えば，$e_x^2(t)$ と $e_w^2(t)$ がほぼ同じ強度であるとき（$S/N = 0$ dB），MTF は $m(f_m) = 0.5$ となる。

2.3 雑音残響環境（$w(t)$ と $h(t)$）における MTF

残響と雑音の両方を考慮した場合の変調伝達関数を考える。前節にならい，図 1 の入出力関係を考えると

$$\begin{aligned} \mathbf{y}(t) &= \int_0^\infty \mathbf{h}(\tau)\mathbf{x}(t-\tau)d\tau + \mathbf{w}(t) \\ e_y^2(t) &= \frac{\overline{e_x^2}}{\alpha}\left[1+\frac{\beta}{\alpha}\cos(2\pi f_m t)\right] + e_w^2(t) \end{aligned}$$

を得る。式 (3) と式 (4) の関係を眺め，前節の導出過程を踏まえると，余弦波にかかる $m(f_m)$ は，次式のように残響・雑音の影響をそれぞれ受けることがわかる。

$$m(f_m) = \frac{1}{\sqrt{1+\left(2\pi f_m\frac{T_R}{13.8}\right)^2}} \times \frac{\overline{e_x^2}}{\overline{e_x^2}+\overline{e_w^2}} \tag{18}$$

以上が，Houtgast と Steeneken によって導入された MTF の概念である。MTF に着目すれば，残響は変調度に関する低域通過フィルタの効果，雑音は変調度のゲインの減衰効果をもたらすことがわかる。本論文では，残響環境下での音声信号処理に焦点を当てるため，以後，残響環境での MTF の概説に特化することにする。

3. パワーエンベロープ逆フィルタ処理

3.1 処理方法

室内音響における入出力関係式（$\mathbf{x}(t)$，$\mathbf{h}(t)$，$\mathbf{y}(t)$）と MTF $m(f_m)$ の対応を再度眺める。2 節ではパワーエンベロープ（$e_x^2(t)$ と $e_y^2(t)$）を単一周波数成分からなるものと仮定したが，これを限定しない形で式 (10) を展開，整理すると，$e_x^2(t)$，$e_h^2(t)$，$e_y^2(t)$ の間には，

$$\begin{aligned}\langle \mathbf{y}^2(t) \rangle &= \left\langle \left\{ \int_{-\infty}^{\infty} \mathbf{x}(\tau)\mathbf{h}(t-\tau)d\tau \right\}^2 \right\rangle \\ &= \int_{-\infty}^{\infty} e_x^2(\tau)e_h^2(t-\tau)d\tau = e_y^2(t) \quad (19)\end{aligned}$$

の関係があることがわかる（詳細な導出過程は付録 3 を参照）。同様に，式 (14) と式 (19) の MTF の関係を眺めると $m(f_m)$ は $e_h^2(t)$ に関するパワーを正規化した一種の Fourier 変換であることもわかる。

以上を踏まえると，MTF に基づく定式化においては，信号だけでなくパワーエンベロープに関しても線形性が成り立っていることがわかる。このことから，仮に残響の効果（$e_h^2(t)$）を事前に知ることができれば，この特性が最小位相特性を有するときに限り，$e_y^2(t)$ から原信号のパワーエンベロープ $e_x^2(t)$ を求めることができる。これは逆畳み込みの関係であることから，パワーエンベロープ逆フィルタ処理と呼ばれる [9]。

次に，これらを計算機で取り扱うために，標本化定理に基づき連続時間信号を離散時間信号に変換する。ここでは，サンプリング周波数 $f_s = 20$ kHz，サンプル点 n とし，$e_x^2[n]$，$e_h^2[n]$，$e_y^2[n]$，$x[n]$，$h[n]$，$y[n]$ と表記する。これら三つのパワーエンベロープの z 変換をそれぞれ，$E_x(z)$，$E_h(z)$，$E_y(z)$ とする。ここで，インパルス応答のパワーエンベロープの z 変換を考える。式 (13) のパワーエンベロープ $e_h^2(t)$ は単純な指数減衰関数であり，$t < 0$ では $e_h^2(t) = 0$ と仮定していることから，最小位相特性を有することがわかる。この特性を踏まえ，$e_h^2[n]$ の z 変換 $E_h(z)$ を求めると

$$E_h(z) = \frac{a^2}{1 - \exp\left(-\frac{13.8}{T_R \cdot f_s}\right) z^{-1}} \quad (20)$$

を得る（導出過程については付録 4 を参照）[2]。式 (19) の畳み込みの関係から，$e_x^2(t)$ の変調スペクトル $E_x(z)$ は $e_y^2(t)$ の変調スペクトル $E_y(z)$ と MTF の逆数（逆フィルタ，IMTF）$1/E_h(z)$ の積で求められる。

$$E_x(z) = \frac{E_y(z)}{a^2} \left\{ 1 - \exp\left(-\frac{13.8}{T_R \cdot f_s}\right) z^{-1} \right\} \quad (21)$$

この逆フィルタは，1 次の IIR フィルタであり，$e_y^2[n]$ に適用することで，最終的に $e_x^2[n]$ を得る [9]。

図 3 は，MTF に基づくパワーエンベロープ逆フィルタ処理を説明するものである。図 3 (a) は，$e_x^2(t) = 0.5(1 + \sin(2\pi f_m t))$ を示す。ここで，f_m は 10 Hz であり，変調度は 1 である。図 3(b) は，式 (5) を利用して得られた原信号 $x(t) = e_x(t) \cdot n_x(n)$ を示す。図 3 (c) は，$T_R = 0.5$ s としたときの式 (13) の $e_h^2(t)$ を示す。図 3 (d) は，残響インパルス応答 $h(t) = e_h(t) \cdot n_h(t)$ を示す。図 3(e) と図 3(f) は，式 (19) のように $e_x^2(t)$ と $e_h^2(t)$ の畳み込みから得られた $e_y^2(t)$ と，$x(t)$ と $h(t)$ の畳み込みから得られた $y(t)$ を示す。図の左側（(a), (c), (e)）はパワーエンベロープを，図の右側（(b), (d), (f)）は各信号波形を示す。この図では，変調度が 1.0（図 3(a)）から 0.404 (図 3(a) の点線と図 3(e) の点線の間の偏差の最大値) に減衰していることがわかる。MTF は，f_m と T_R の関数とした変調度 [6] を示すため，式 (14) に $f_m = 10$ Hz と $T_R = 0.5$ s を代入することで，$m(f_m) = 0.402$ を得る。図 3 (g) の実線は，$T_R = 0.5$ として，式 (21) を利用した際の，$e_y^2(t)$（図 3 (e)）から求めた $\hat{e}_x^2(t)$ を示す。この結果から，パワーエンベロープ逆フィルタが，パワーエンベロープの大きさと形状に関して，正確に回復可能であることを示している。

3.2 応用上の問題点

上記で概説したパワーエンベロープ逆フィルタの原理は，広林ら [9] によって最初に提案されたが，次に述べるような原理上の問題点が多々残されていた。

図 4 は，パワーエンベロープ逆フィルタ処理のブロックダイアグラムを示す。このモデルの概念において，観測された残響信号 $y(t)$ から完全に正確にパワーエンベロープ $e_y^2(t)$ を抽出でき，かつ室内インパルス応答のパラメータ（振幅項 a と残響時間 T_R）を事前に知っているという条件であれば，図 4 の方法は観測された残響信号 $y(t)$ から原信号のパワーエンベロープを正確に回復することができる。しかし，$e_y^2(t)$ が $y(t)$ から正確に抽出されなかった場合，この方法は式 (21) の逆フィルタ法を利用して $y(t)$ から $e_x^2(t)$ を正確に回復できない。更に，$e_y^2(t)$ が $y(t)$ から完全に正確に抽出されたとしても，二つのパラメータ（a と T_R）の値が不正確であれば，この方法は正しく $\hat{e}_x^2(t)$ を回復できない。

例えば，図 3 (g) の実線とは対照的に，逆フィルタが不適切なパラメータ値（例えば，$T_R = 0.3$ s や $T_R = 1.0$ s）で適用された場合，図 3 (g) の破線，一点鎖線に示すように，$e_x^2(t)$ は正確に回復されない。ここで，前者の場合，“過小回復” であり，後者の場合，“過剰回復” となる。特に，$T_R = 1.0$ s（図 3 (g) の破線）のとき，$\hat{e}_x^2(t)$ は $e_y^2(t)$ から過剰に処理され，これは負の値をもつことになる（物理的にパワーが負になることは有り得ない）。ここで，$\hat{e}_x(t)^2$ の改善度は，$T_R = 0.3$ s（3 (g) の点線）のときよりも高くなり，$T_R = 0.5$ s での回復より低くなる。これは，式 (19) が積分処理であるのに対し，式 (21) が微分処理であることに起因している。

以上のことから，パワーエンベロープ逆フィルタ処理を利用するには，次の二つの問題点を解決しなければならない。(1) パワーエンベロープをどのようにして正確に抽出するか，(2) 室内インパルス応答のパラメータ（振幅項 a と残響時間 T_R）をどのように決定するか。これらの解決法について，次節で概説する。

[2] パワーエンベロープ畳み込みの他，エンベロープを 2 乗しない形での展開も議論されている [18, 19]。

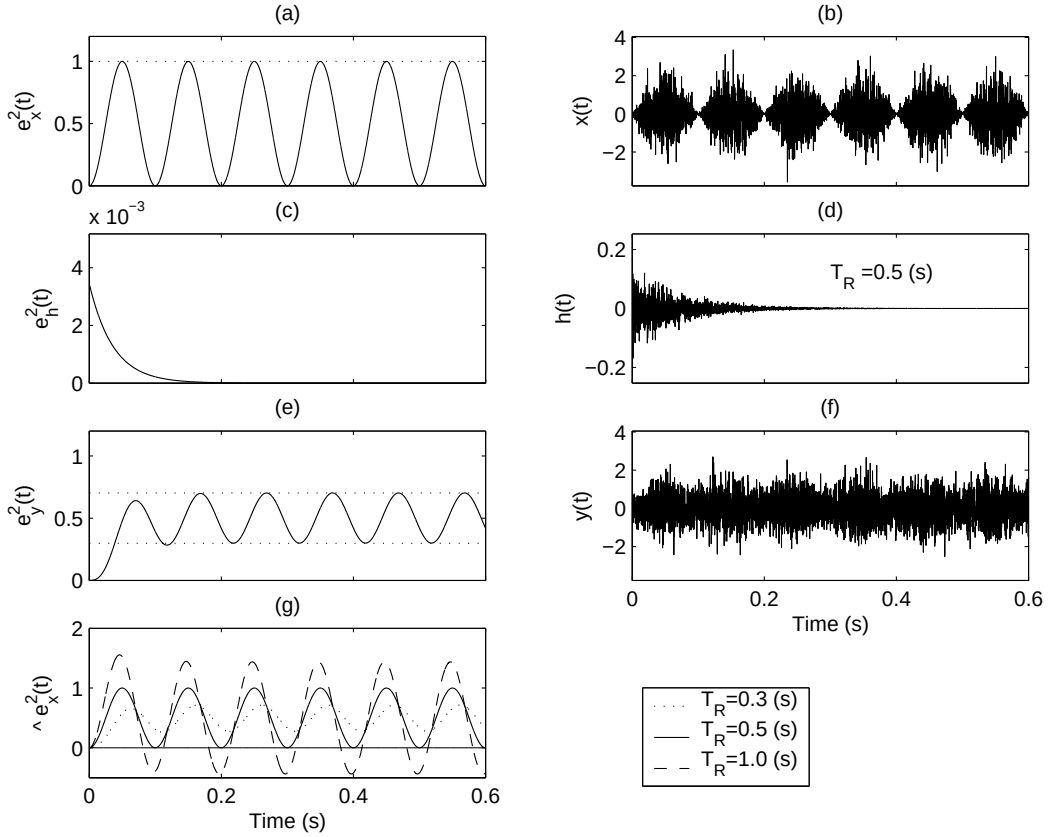

図 3　MTF の概念に基づいた伝達系のパワーエンベロープ間の関係。(b) 入力信号 $x(t)$ と (a) そのパワーエンベロープ $e_x^2(t)$，(d) 室内音響インパルス応答 $h(t)$ と (c) そのパワーエンベロープ $e_h^2(t)$，(e) $e_x^2(t) * e_h^2(t)$ から得られたパワーエンベロープ $e_y^2(t)$，(f) $x(t) * h(t)$ から得られた出力信号 $y(t)$，(g) 回復されたパワーエンベロープ $\hat{e}_x^2(t)$

Fig. 3 Example of the relationship between the power envelopes of a system based on the MTF concept: (a) power envelope $e_x^2(t)$ of (b) original signal $x(t)$, (c) power envelope $e_h^2(t)$ of (d) impulse response $h(t)$, (e) power envelope $e_y(t)^2$ derived from $e_x^2(t) * e_h^2(t)$, (f) reverberant signal $y(t)$ derived from $x(t) * h(t)$, and (g) restored power envelope $\hat{e}_x^2(t)$

3.3　改良法

3.3.1　パワーエンベロープ抽出法

パワーエンベロープ抽出法に関しては，信号伝送処理で扱われるような振幅変調信号に対する復調処理を考えることで，適切な解法を得ることができる。しかし，MTF に基づくパワーエンベロープ逆フィルタ処理では，キャリアが正弦波ではなく白色雑音であるため，一般的な復調方式（例えば，半波整流法 [20] など）では対応できない。これに対し，著者の研究グループによって二つの抽出法が提案された [10]。一つは，式 (9) の集合平均を利用する方法である[3]。もう一つの方法は，Hilbert 変換の関係 [21] を利用したものである。両手法とも同程度の抽出精度を有するが，計算が簡単で計算量が低いことから，本稿では後者を紹介する。

[3]式 (9) では，キャリアが白色ガウス過程の確率変数としているため，実際の観測信号 $y(t)$ から集合平均を求めることができない。白色ガウス過程に基づいて生起された白色雑音は互いに無相関であることから，相異なる白色雑音を掛け合わせたものも白色雑音であると考える。提案法は，この仮定に基づき，有限個の白色雑音を観測信号に掛け合わせて得た擬似的な観測集合 $\mathbf{y}(t)$ の集合平均を求めることで，$e_y^2(t)$ を得るものである。

Hilbert 変換器は，90 度位相器と呼ばれ，偶関数・奇関数の変換や解析信号からの瞬時振幅の算出によく利用される [21]。観測信号 $y(t)$ のキャリアが偶関数あるいは奇関数で構成されていると仮定すれば，Hilbert 変換を利用して瞬時振幅を求め，これに低域通過フィルタ処理（LPF）をすることでエンベロープを抽出することができる。この方法を次に示す。

$$\hat{e}_y^2(t) = \mathbf{LPF}\left[\left|y(t) + j \cdot \mathbf{Hilbert}(y(t))\right|^2\right] \qquad (22)$$

ここで，LPF のカットオフ周波数は 20 Hz とした。これは音声の主要な変調周波数が 1〜16 Hz にあるという報告に基づいたためである [10].

3.3.2　インパルス応答のパラメータ推定

パワーエンベロープ逆フィルタ法では，残響時間 T_R と振幅項 a が逆フィルタを構成する主要素となるため，残響回復の前にこれらの正確な値を知らなければならない。著者の研究グループでは，次のような考えに沿って，パラメータ推定法を提案した [10]。

まず，図 2 に示すように，パワーエンベロープの変調周波数 f_m が単一の場合を考える。先に述べたように，

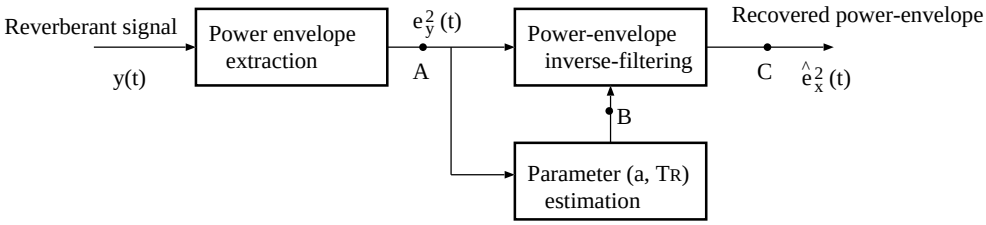

図 4　パワーエンベロープ逆フィルタ処理のブロックダイアグラム
Fig. 4 Block diagram of the power envelope inverse filtering method

f_m と $m(f_m)$ の関係は，残響による $e_x^2(t)$ の変調度の減衰量を表しているのだから，$1/m(f_m)$ 倍（IMTF）だけ $e_y^2(t)$ の変調度を補正することが，残響回復となる。そのため，完全な回復量を示す T_R を探し出すことで，MTF を形成したときの T_R を知ることができる。この算出は，f_m が単一周波数のときに容易であるが，複数周波数のときは難しい。そこで，IMTF の算出を変調周波数領域ではなく，時間領域で行うことを考える。図 2(g) に示したように，パワーエンベロープの過小・過剰処理の結果から，残響を付与したときの残響時間と同じ T_R の値を利用してパワーエンベロープを回復すると，$\hat{e}_x^2(t)$ が $e_x^2(t)$ と一致することがわかる。

パワーエンベロープは少なくとも一つの 0 の値を取るディップないし無音区間をもつため，ここでは，$e_x^2(t)$ の変調度が 1 であると仮定する。上記の発見に基づき，オリジナルと回復されたパワーエンベロープの間の一致条件が，残響によって減少した変調度を回復することであると定義する。この条件は，ちょうど T_R を推定する条件となっており，次式のように，回復されたパワーエンベロープ $\hat{e}_x^2(t)$ の最大のディップが 0 であるところ，あるいは $\hat{e}_x^2(t)$ の負の面積がちょうど 0 であるところを検出することで推定することができる [10]。

$$\hat{T}_R = \max\left(\underset{T_{R,\min}\le T_R\le T_{R,\max}}{\arg\min}\int_0^T \left|\min\left(\hat{e}_{x,T_R}^2(t),0\right)\right| dt\right) \tag{23}$$

但し，T は $y(t)$ の信号長，$\hat{e}_{x,T_R}^2(t)$ は T_R を変数として得られたパワーエンベロープである。“$\max(\arg\min\{\cdot\})$” は，これらの候補 $\hat{e}_{x,T_R}^2(t)$ の中から，負の面積がちょうど 0 になるときの T_R の最大値を求めることを意味する。そのため，この式は，回復されたパワーエンベロープがちょうど負の値をもつ前のところで制約を受けていることを意味する。ここで，$T_{R,\min}$ と $T_{R,\max}$ はそれぞれ T_R の取り得る範囲の下限値と上限値を表す。

例えば，図 2(g) に $T_R = 0.3$, 0.5, 1.0 s のときに回復した $\hat{e}_{x,T_R}^2(t)$ の候補を示す。ここで，$T_{R,\min} = 0.0$ s，$T_{R,\max} = 3.0$ s とした。 $T_R = 0.3$ s のときの候補 $\hat{e}_{x,T_R}^2(t)$ はちょうど過小回復の状態にあり，$T_R = 1.0$ s のときの候補 $\hat{e}_{x,T_R}^2(t)$ は過剰回復の状態にある。式 (23) を利用して T_R を推定すると，最適値 $\hat{T}_R = 0.5$ s が三つの候補の中で最適であることがわかる。

次に振幅項 a は，室内インパルス応答の増幅度に関係する。式 (13) と式 (21) では， 同じ値を利用したが，実際，室内における残響効果は信号の増幅よりも反射による遅延の重ね合わせの効果が強い。そこで，$e_h^2(t)$ の伝達特性と MTF の整合をとるために次式のようにパワーによる正規化を行うことにした。

$$\hat{a} = \sqrt{1/\int_0^T \exp\left(-\frac{13.8t}{\hat{T}_R}\right) dt} \tag{24}$$

実環境における応用では，この値を推定しなければいけないかもしれないが，室内インパルス応答のパワーエンベロープから最適化手法により求めることもできる。

4. 評価シミュレーション

ここでは，著者の研究グループによって改良されたパワーエンベロープ逆フィルタ法 [10] を評価する。そこで，次に示す三つのパワーエンベロープと白色雑音の積で構成される原信号 $x(t)$ を利用する。

1. 正弦波: $e_x^2(t) = 1 - \cos(2\pi F t)$
2. 調波複合音: $e_x^2(t) = 1 + \dfrac{1}{K}\displaystyle\sum_{k=1}^{K}\sin(2\pi k F_0 t + \theta_k)$
3. 帯域制限雑音: $e_x^2(t) = \mathbf{LPF}[n_w(t)]$

但し，$F = 10$ Hz, $F_0 = 1$ Hz, $K = 20$, θ_k はランダム位相，$\mathbf{LPF}[\cdot]$ のカットオフ周波数は 20 Hz とした。ここでは，パワーエンベロープ抽出精度とパワーエンベロープの回復精度の両方を同時に評価する。評価シミュレーションでは，三つのパワーエンベロープに対し，100 個のキャリア（白色雑音）を乗じて得た音源信号 $x(t)$ および，５種類の残響時間（$T_R = 0.1, 0.3, 0.5, 1.0, 2.0$ s）に対して，それぞれ 100 種類の室内インパルス応答 $h(t)$ を畳み込んで得られた合計 $1,500$（$= 3\times 5\times 100$）個の残響信号 $y(t)$ を用意する。評価尺度は，パワーエンベロープに対する (1) 相関値と (2) SNR（S をオリジナル，N をオリジナルと推定したものとの差）とした。

$$\begin{aligned}&\mathrm{Corr}(e_x^2, \hat{e}_x^2)\\ &= \frac{\int_0^T \left(e_x^2(t) - \overline{e_x^2(t)}\right)\left(\hat{e}_x^2(t) - \overline{\hat{e}_x^2(t)}\right) dt}{\sqrt{\left\{\int_0^T (e_x^2(t) - \overline{e_x^2(t)})^2 dt\right\}\left\{\int_0^T (\hat{e}_x^2(t) - \overline{\hat{e}_x^2(t)})^2 dt\right\}}}\end{aligned} \tag{25}$$

$$\begin{aligned}&\mathrm{SNR}(e_x^2, \hat{e}_x^2)\\ &= 10\log_{10}\frac{\int_0^T (e_x^2(t))^2 dt}{\int_0^T (e_x^2(t) - \hat{e}_x^2(t))^2 dt} \quad \text{(dB)}\end{aligned} \tag{26}$$

但し，$\overline{e_x^2(t)}$ は $e_x^2(t)$ の平均値を表す。

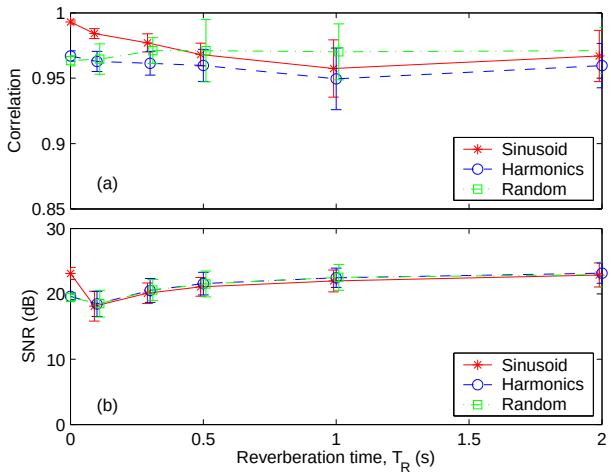

図 5　パワーエンベロープの抽出精度：(a) 相関値，(b) SNR

Fig. 5 Extraction accuracy of the power envelope: (a) correlation, (b) SNR

4.1　パワーエンベロープ抽出精度の評価

はじめに，パワーエンベロープ抽出法の精度について評価する（図 4 の A 点での評価）。図 5 に，$x(t)$ ないし，$y(t)$ から抽出された三つのパワーエンベロープの抽出精度を示す。これは式 (22) を利用して得られた結果であり，図 5(a) は相関値を，図 5(b) は SNR を示す。$e_x^2(t)$ の抽出精度については，オリジナルとの差を，$e_y^2(t)$ の抽出精度については，式 (19) から得られたパワーエンベロープとの差を評価している。$T_R = 0$ s のときは，$e_x^2(t)$ の評価に相当する。この結果から，残響時間にほとんど影響を受けずに，高い精度でパワーエンベロープを $y(t)$ から抽出できることが確認された。

4.2　残響時間 T_R の推定精度の評価

次に，式 (23) を利用した残響時間の推定精度を評価する（図 4 の B 点での評価）。図 6(a) は，各エンベロープに対して推定された残響時間 $\hat{T}_R$ を示す。各点は，各 100 個の観測信号から残響時間を推定したときの平均値を示し，エラーバーはそのときの標準偏差を示す。図中の点線は，オリジナルの残響時間 T_R を示す。残響時間が 0.5 s 付近までは，推定値がオリジナルの T_R とほぼ等しくなっているが，それ以降は徐々に差が生じる傾向にある。この影響については，次節で説明する。

4.3　パワーエンベロープ回復精度の評価

最後に，図 5 と図 6(a) の結果を踏まえ，式 (21) から求められた $\hat{e}_x^2(t)$ の回復精度を評価する（図 4 の C 点での評価）。ここでは，回復精度として，相関値の改善度 $\mathrm{Corr}(e_x^2, \hat{e}_x^2) - \mathrm{Corr}(e_x^2, e_y^2)$ と，SNR の改善度 $\mathrm{SNR}(e_x^2, \hat{e}_x^2) - \mathrm{SNR}(e_x^2, e_y^2)$ を評価した。図 3(a) と図 3(e) に示すように，各エンベロープの変調度の関係 (MTF) は，T_R が増加するにつれ減衰する。同様に，$\mathrm{Corr}(e_x^2, e_y^2)$ と $\mathrm{SNR}(e_x^2, e_y^2)$ も，T_R の増加とともに減衰する。そのため，$\hat{e}_x^2(t)$ が正しく回復されたのであれば，相関値と SNR の改善度は，両方とも正の値を示すはずである。もしいずれかの値が負の値を示すのであれば，$\hat{e}_x^2(t)$ が適切に回復されたとはいえない。

図 6 に，パワーエンベロープ逆フィルタ処理による回復精度を示す。図 6(b), (d), (f) は，それぞれ，三つのパワーエンベロープに対する相関値の改善度を，図 6(c), (e), (g) は SNR の改善度を示す。図中の実線は，式 (23) を利用して T_R を推定して $\hat{e}_x^2(t)$ を求めたときの結果を，破線は T_R を既知として $\hat{e}_x^2(t)$ を求めたときの結果を示す。各点とエラーバーは，それぞれ結果の平均値と標準偏差を示す。図 6(b)-(g) の改善度は，すべての条件で正の値を示していることから，パワーエンベロープ逆フィルタ法が，残響によって歪んだパワーエンベロープを非常に効果的に回復しているといえる。

図中の実線と破線の改善度の違いについて検討する。図 6 (b)-(c) の T_R が 0.3〜1.0 s のときの結果と図 6(d)-(g) の T_R が 0.5〜2.0 s のときの結果を除き，二つの改善度には大きな違いが見られない。逆に，ここであげた違いが見られる場所は，図 6(a) に示す T_R の推定精度が低下しているところと関係している。残響時間が比較的短い場合（$T_R < 0.5$），推定値 $\hat{T}_R$ は非常に点線に近い値となっており，相関値と SNR の改善度も非常によい値になっている。しかし，T_R が 0.5 を超えたあたりから，その推定精度は低下し，SNR の改善度も正の値を示しているが若干低下する傾向にある。相関値の改善度は逆に増加するか同じ値を保持していることから，T_R 推定の違いが回復されたパワーエンベロープの形状には大きな影響を与えていないということがわかる。SNR の改善度が低下しているのは，T_R の過小推定が振幅項 $\hat{a}$ に影響を与えているためと考えられる。比較的長い残響時間（$T_R > 0.5$）では，T_R を既知とした場合よりも，T_R を推定したほうがよい改善度を示している。これは，図 6(a) の T_R 推定の結果が悪いのではなく，図 6(b)-(g) に示す回復精度のレベルで最良値を取るような制約を式 (23) が表しているといえる。

上記について，事例を通して検討する。図 7 に，$T_R = 1.0$ s のときの正弦波ならびに帯域制限雑音のエンベロープに対する回復結果の一例を示す。図 7(a) と図 7(c) の $e_y^2(t)$ は，$y(t)$ から抽出されたものである（破線）。図中の 1 点鎖線と実線は，それぞれ，$e_x^2(t)$ と式 (19) を利用して得られた $e_y^2(t)$ を示す。図 7(b) と図 7(d) で示される負の面積は，パワーエンベロープのディップの最大値付近で検出される（図 7(b) の場合，0.1 s 付近；図 7(d) の場合，0.2 s 付近）。ここで，$\hat{T}_R$ は図 7(b) のとき 0.73 s，図 7(d) のとき 0.78 s であった。図 7(a) の $e_x^2(t)$ と $e_y^2(t)$ の間の相関値と SNR は 0.15 と 4.44 dB であったのに対し，図 7(b) の $e_x^2(t)$ と $\hat{e}_x^2(t)$ の間の相関値と SNR はそれぞれ 0.98 と 15.09 dB であった。これらの結果から，相関値と SNR の改善度は，それぞれ，0.83 と 10.65 dB であった。対照的に，T_R を既知としてパワーエンベロープを回復したとき，その改善度は，相関値で 0.82（$= 0.97 - 0.15$），SNR で 7.59（$= 12.03 - 4.44$）dB であった。そのため，T_R を既知として利用するよりも推定した T_R を利用したほうが，適切な改善度が得られることがわかる。このときの効果は，相関値で

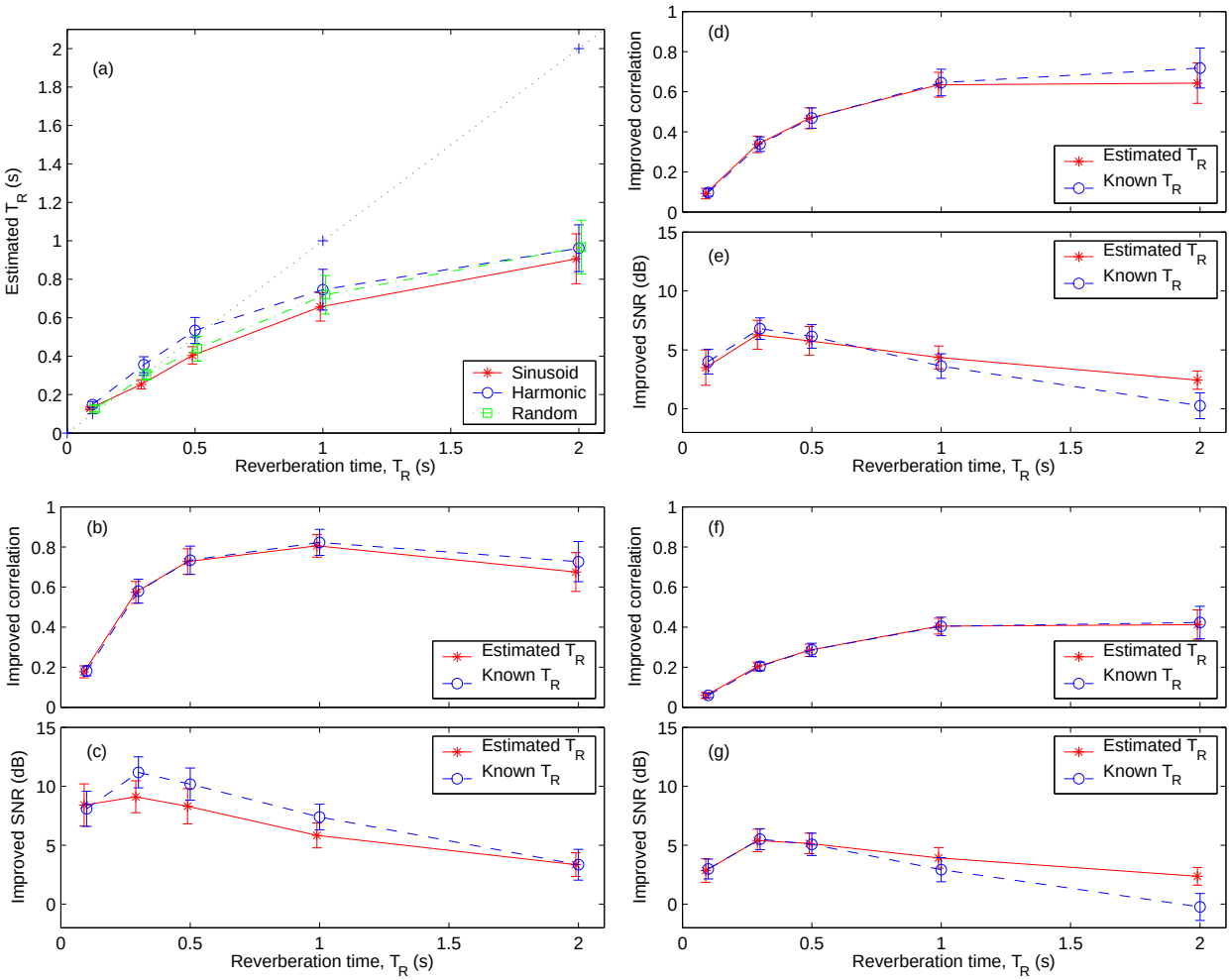

図 6　評価シミュレーションの結果。(a) 残響時間の推定値（点線は理想化された残響時間）。正弦波（(b) と (c)），調波複合音（(d) と (e)），帯域制限雑音（(f) と (g)）のパワーエンベロープに対する回復精度の比較（相関値の改善度と SNR の改善度）

Fig. 6 Results of simulations for evaluation: (a)Estimated reverberation time. The dotted line shows the idealized reverberation time. Comparisons with the envelope restoration accuracy for sinusoidal power envelope ((b) improved correlation and (c) improved SNR), a harmonic power envelope ((d) improved correlation and (e) improved SNR), and a band-limited noise power envelope: (f) improved correlation and (g) improved SNR

0.01，SNR で 3.06 dB であった。

図 7(d) に，帯域制限雑音をパワーエンベロープとした場合に対する同様の評価を示す。ここでは，T_R を既知とした場合，相関値で 0.57 （$= 0.97 - 0.40$），SNR で 6.44 dB（$= 12.8 - 6.36$） dB であったのに対し，T_R を推定してエンベロープを回復した場合，相関値で 0.58（$= 0.98 - 0.40$），SNR で 8.04 dB（$= 14.4 - 6.36$）であった。この結果， T_R を推定したことによる効果は，相関値で 0.01，SNR で 1.60 dB であった。

5.　おわりに

本稿では，変調伝達関数に基づく音声信号処理（全３回シリーズ）の第１稿として，Hougast と Steeneken によって導入された変調伝達関数の考えと，変調伝達関数に基づいたパワーエンベロープ逆フィルタ処理の原理を概説した。また，この原理を実際の信号処理に適用するときの問題点と改善法を紹介した。次回（第 2 稿）では，本稿で紹介したパワーエンベロープ逆フィルタ処理を，残響音声回復に応用した手法を概説する。

謝　辞

本論文で紹介する研究は，科学研究費補助金若手研究 B（No.14780267），若手研究 A（No. 18680017），萌芽研究（No. 17650048），科学技術振興調整費（若手研究支援プログラム），矢崎科学技術振興記念財団（特定研究助成）ならびに総務省 戦略的情報通信研究開発推進制度（課題番号 071705001）の援助を受けて行われた。研究協力者である，本学 赤木正人教授，Lu Xugang 博士，本学修了生の古川正和君，酒田恵吾君，戸井真智君，柴野洋平君，細呂木谷敏弘君，平松壮太君，本学在学生の山崎 悠君に心より感謝する。

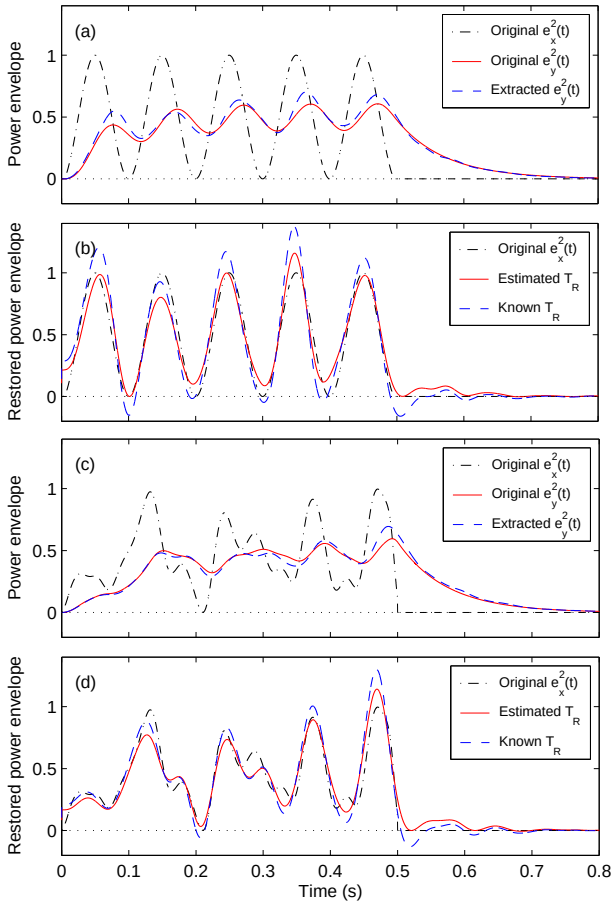

図 7　結果例：(a) 正弦波信号のパワーエンベロープと (b) $T_R = 1.0$ s のときに回復されたパワーエンベロープ，(c) 帯域制限雑音のパワーエンベロープと (d) $T_R = 1.0$ s のときに回復されたパワーエンベロープ

Fig. 7 Sample results: (a) extracted power envelope and (b) restored power envelope for sinusoidal with $T_R = 1.0$ s; (c) extracted power envelope and (d) restored power envelope for a band-limited noise with $T_R = 1.0$ s

参考文献

[1] A.S. Bregman: Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press, Cambridge, MA, 1990.

[2] 戸井田 義徳: 小特集–音声の明瞭度と認識率–, 空間内における音声情報伝達, 音響誌, Vol. 51, No. 4, pp. 312–316, 1995.

[3] N. R. French and J. C. Steinberg: Factors governing the intelligibility of speech sounds, J. Acoust. Soc. Am., Vol. 19, pp. 90–119, 1947.

[4] T. Houtgast and H. J. M. Steeneken: The modulation transfer function in room acoustics as a predictor of speech intelligibility, Acustica, Vol. 28, pp. 66–73, 1973.

[5] T. Houtgast, H. J. M. Steeneken, and R. Plomp: Predicting speech intelligibility in rooms from the Modulation Transfer Function. I. General Room Acoustics, Acustica, Vol. 46, pp. 60–72, 1980.

[6] 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, pp. 1069–1077, 1985.

[7] 小椋靖夫，浜田晴夫，三浦種敏: 音場における音声伝達品質のための MTF と STI, 音響誌, Vol. 40, pp. 181–191, 1984.

[8] H. Kuttruff: Room Acoustics, Taylor & Francis, fourth edition, London, 2000.

[9] 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, 信学論 A, Vol. J81-A, No. 10, pp. 1323–1330, 1998.

[10] 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, Acoustical Science and Technology, Vol. 25, No. 4, pp. 232–242, 2004.

[11] M. Unoki, K. Sakata, M. Furukawa, and M. Akagi: A speech dereverberation method based on the MTF concept in power envelope, Acoustical Science and Technology, Vol. 25, No. 4, pp. 243–254, 2004.

[12] M. Unoki and T. Hosorogiya: Estimation of fundamental frequency of reverberant speech by utilizing complex cepstrum analysis, J. Signal Processing, Vol. 12, No. 1, pp. 31–44, Jan. 2008.

[13] S. Hiramatsu and M. Unoki: A Study on the Blind Estimation of Reverberation Time in Room Acoustics, J. Signal Processing, Vol. 12, No. 4, pp. 323–326, July 2008.

[14] S. Irie and S. Hirobayashi: Simulation of an acoustic system using power envelope inverse filtering, Proc. ICASSP2008, pp. 325–328, Las Vegas, 2008.

[15] A. Papouris: Probability: Random Variables, and Stochastic Processes, Third Edition, MacGraw-Hill, Inc., New York, 1991.

[16] M. R. Schroeder: Modulation Transfer Functions: Definition and Measurement, Acustica, Vol. 49, pp. 179–182, 1981.

[17] M. R. Schroeder: New Method of Measuring Reverberation Time, J. Acoust. Soc. Am., Vol. 37, No. 6, pp. 409, 1965.

[18] 広林茂樹, 寺島洋行，山淵龍夫: 残響音場における音響信号のエンベロープ推定法の評価, シミュレーション, Vol. 22, No. 3, pp. 68–75 (208–215), 2003.

[19] X. Lu, M. Unoki, and M. Akagi: A study on restoration methods for temporal envelope of reverberant speech based on the modulation transfer function concept, Proc. JCA2007, P-2-11, Sendai, Japan, 2007.

[20] F. J. Taylor: Principles of Signals and Systems, MacGraw-Hill, Inc., New York, 1994.

[21] A. V. Oppenheim and R. W. Schafer: Digital Signal Processing, Prentice-Hall, Inc., London, 1975.

付録 1：式 (8) と式 (9) の導出

$\mathbf{v}(t)$ をランダム変数，$f(v,t)$ を $\mathbf{v}(t)$ の確率密度関数とする。このとき，集合平均は次式で定義される [15]。

$$< \mathbf{v}(t) >= \int_{-\infty}^{\infty} v f(v,t) dv \qquad (27)$$

式 (7) の関係に注意して，式 (8) の左辺を求めると

$$\begin{aligned} < \mathbf{x}^2(t) > &= < e_x^2(t)\mathbf{n}_x^2(t) > \\ &= e_x^2(t) < \mathbf{n}_x^2(t) > \\ &= e_x^2(t) \end{aligned}$$

を得る。但し，$< \mathbf{n}_x^2(t) >= \delta(0) = 1$ である。式 (9) の $< \mathbf{y}^2(t) >$ についても同様の導出過程から得られる。

付録 2：式 (11) の導出

式 (10) の $< \mathbf{y}^2(t) >$ は $\mathbf{x}(t)$ と $h(t)$ の畳み込みを 2 乗した集合平均と等しいため，次式となる。

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

これを整理すると

$$\begin{aligned} &< \mathbf{y}^2(t) > \\ &= \int_0^{\infty} h^2(\tau)e_x^2(t-\tau) < \mathbf{n}_x(t-\tau)\mathbf{n}_x(t-\tau) > d\tau \\ &= \int_0^{\infty} h^2(\tau)(\overline{e_x^2}(1+\cos(\omega_m(t-\tau)))d\tau \\ &= \overline{e_x^2}\left[\int_0^{\infty} h^2(\tau)d\tau + \int_0^{\infty} h^2(\tau)\cos(\omega_m(t-\tau)))d\tau\right] \\ &= \overline{e_x^2}\left[\int_0^{\infty} h^2(\tau)d\tau + \mathcal{R}\left\{\int_0^{\infty} h^2(\tau)e^{j\omega_m(t-\tau)}d\tau\right\}\right] \\ &= \overline{e_x^2}\left[\int_0^{\infty} h^2(\tau)d\tau \right. \\ &\qquad \left. +\mathcal{R}\left\{e^{j2\pi f_m t}\int_0^{\infty} h^2(\tau)e^{-j2\pi f_m \tau}d\tau\right\}\right] \\ &= \frac{\overline{e_x^2}}{\alpha}\left[1+\frac{\beta}{\alpha}\cos(2\pi f_m t)\right] \end{aligned}$$

を得る。但し，$\mathcal{R}$ は実数部を意味し，$\alpha = \int_0^{\infty} h^2(\tau)d\tau$，$\beta = \int_0^{\infty} h^2(\tau)\exp(-j\omega_m\tau)d\tau$ である。ここで，式 (4) との対応を見れば，$m(f_m) = \beta/\alpha$ から，式 (12) を得る。

付録 3：式 (19) の導出

式 (19) の $< \mathbf{y}^2(t) >$ は $\mathbf{x}(t)$ と $\mathbf{h}(t)$ の畳み込みを 2 乗した集合平均と等しいため，次式となる。

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

ここで，式 (5) と式 (13) を代入し，整理すると

$$\begin{aligned} &< \mathbf{y}^2(t) > \\ &= \left\langle \int_{-\infty}^{\infty} e_x(\tau_1)\mathbf{n}_x(\tau_1)e_h(t-\tau_1)\mathbf{n}_h(t-\tau_1)d\tau_1 \right. \\ &\qquad \left. \times \int_{-\infty}^{\infty} e_x(\tau_2)\mathbf{n}_x(\tau_2)e_h(t-\tau_2)\mathbf{n}_h(t-\tau_2)d\tau_2 \right\rangle \\ &= \int_{-\infty}^{\infty} e_x(\tau_1)e_h(t-\tau_1)\int_{-\infty}^{\infty} e_h(\tau_2)e_h(t-\tau_2) \\ &\qquad \times < \mathbf{n}_x(\tau_1)\mathbf{n}_x(\tau_2) > \\ &\qquad \times < \mathbf{n}_h(t-\tau_1)\mathbf{n}_h(t-\tau_2) > d\tau_1 d\tau_2 \end{aligned} \tag{28}$$

を得る。但し，$\mathbf{n}_x(t)$ と $\mathbf{n}_h(t)$ は，平均 0，分散 1 を有するランダム関数であり，式 (7) の関係をもつ。ここで，$\tau = \tau_1 = \tau_2$ と式 (9) を代入すると，次式を得る。

$$\begin{aligned} &< \mathbf{y}^2(t) > \\ &= \int_{-\infty}^{\infty} e_x(\tau_1)e_h(t-\tau_1)\int_{-\infty}^{\infty} e_x(\tau_2)e_h(t-\tau_2) \\ &\qquad \times\delta(\tau_2-\tau_1)^2 d\tau_1 d\tau_2 \\ &= \int_{-\infty}^{\infty} e_x^2(\tau)e_h^2(t-\tau)d\tau \\ &= e_x^2(t) * e_h^2(t) = e_y^2(t) \end{aligned}$$

付録 4：式 (21) の導出

パワーエンベロープ $e_h^2[n]$ の z 変換は

$$E_h(z) = \sum_{n=-\infty}^{\infty} e_h^2[n]z^{-n}$$

である。但し，$t = n/f_s$ であり，$t < 0$ で $e_h^2(t) = 0$ であるため，n の下限は 0 である。これを整理すると，

$$\begin{aligned} E_h(z) &= \sum_{n=0}^{\infty} a^2 \exp\left(-\frac{13.8n}{f_s \cdot T_R}\right) z^{-n} \\ &= a^2\gamma^n z^{-n} \\ &= a^2 + a^2\gamma z^{-1} + a^2\gamma^2 z^{-2} + a^2\gamma^3 z^{-3} + \cdots \\ &= \frac{a^2}{1-\gamma z^{-1}} \\ &= \frac{a^2}{1-\exp\left(-\frac{13.8}{f_s \cdot T_R}\right) z^{-1}} \end{aligned}$$

を得る。但し，$\gamma = \exp(-13.8/f_s \cdot T_R)$ である。

鵜木 祐史 1994 年職業能力開発大学校情報工学科卒。1996 年北陸先端科学技術大学院大学情報科学研究科博士前期課程修了。1999 年同大情報科学研究科博士後期課程修了。博士（情報科学）。同年 ATR 人間情報通信研究所第一研究室客員研究員，2000 年英国ケンブリッジ大学生理学部 CNBH 客員研究員，2001 年北陸先端科学技術大学院大学情報科学研究科助手を経て，2005 年同大助教授に奉職（2007 年に准教授）。現在に至る。1998 年〜2001 年の間，日本学術振興会特別研究員（DC2, PD の 2 期）を兼任。主に，聴覚機能のモデル化とそれに基づく信号処理ならびに音声信号処理（残響音声回復，骨導音声回復，音響電子透かし）の研究に従事。日本音響学会佐藤論文賞（1999 年度）ならびに山下太郎学術奨励賞（2005 年度）受賞。信号処理学会，日本電子情報通信学会，日本音響学会，アメリカ音響学会，IEEE，ISCA，EURASHIP 各会員。