# 変調伝達関数に基づく音声信号処理 (2)
## —ブラインド残響音声回復法—
# Speech Signal Processing Based on the Concept of Modulation Transfer Function (2)
## — Blind Speech Dereverberation Method —

鵜木祐史
Masashi Unoki

### 1. はじめに

本論文（全3回シリーズ）の第1稿 [1] では，Houtgast と Steeneken が示した変調伝達関数（MTF）の概念を解説し，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した。本稿（第2稿）では，この応用例として，MTF に基づいたブラインド残響音声回復法を概説し，どの程度のことが現在までにできているのか，何が残された課題であるかを説明する。ここでは，手法の評価として，パワーエンベロープの回復精度だけではなく，残響によって低下した音声明瞭度の改善や音声認識率の改善効果も紹介する。

### 2. 残響音声回復法の取組み

残響の影響を受けた信号から元の音源信号を回復する研究は，音声認識や拡声会議通話，補聴システムといった音声信号処理において重要な課題である。そのため，室内音場で観測された残響音声から原信号を回復する逆フィルタ処理が数多く検討されてきた。

例えば，Neely と Allen は，単一マイクロホンで受音された信号から室内伝達特性の最小位相成分のみを取り除いて残響信号を回復させる方法を提案した [2]。しかし，この方法は，室内音場が最小位相特性を持つときにしか有効ではない。これに対し，三好と金田は，音源の数に対しマイクロホンを1つ以上多く配置することで，音源とマイクロホンの間の伝達特性の零点が重複しない場合，系が最小位相特性を有してい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 [3]（MINT 法[1]）を提唱した。また，王と板倉は，それぞれの帯域毎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た [6]。室内伝達特性の逆特性を逆フィルタ処理で実現するこれらの方法では，事前に系の伝達特性を測定しておかなければならない。しかし，この伝達特性は，室内環境の様々な変動（例えば，室内の形状や人などの物体の移動，室温変化など）に伴い時々刻々と変動するため，高い回復精度を保持するためには，その都度，伝達特性を測定し直す必要がある[2]。様々な音声信号処理への応用を考えた場合，これらの方法は，瞬時的な系の測定を必要とする点で現実的ではない。

最近，調波構造に基づく音声信号のブラインド残響除去法が，中谷と三好の研究グループによって提案されている [7, 8]。この方法では，室内のインパルス応答を測定する必要がない代わりに，調波構造を定める際，残響音声からの正確な基本周波数（$F_0$）の推定を必要とする。この $F_0$ の推定精度に起因した回復精度の低下が危惧されるが，彼らの報告 [9] では，音声品質の明瞭性の客観的評価と音声認識性能の改善法が検討され，大きな改善効果が得られていることから，有効な手法の一つとして期待されている。

一方，これまでの方法とは別のアプローチとして，変調伝達関数（MTF）の概念に基づいた残響除去法が提案されている[3]。ここでは，室内のインパルス応答を測定することなく，観測した残響音声のエンベロープ情報あるいは変調スペクトルを回復させる方法が提案されている。例えば，Langhans と Strube[10] や Avendano と Hermansky[11] の手法である。これらは，MTF の概念に基づき，短時間 Fourier 変換（STFT）上のパワースペクトル表現とこれの高域通過処理を用いて，室内伝達特性のエンベロープの逆特性を畳み込むというものである。この他に，Mourjopoulos と Hammond[12] や広林ら [13] の手法もある。これらは，STFF ではなく，帯域通過フィルタバンクを用いて残響音声のエンベロープを回復する処理を実現している。これらの手法の違いは，振幅変調の表現に基づいたとき，時間的·

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

[1] 一意な解を求めるための逆行列演算に膨大な時間を要することが問題であったが，現在では，MINT 法を発展させた最小ノルム解に基づく逆フィルタ理論 [4] や対象音源に制約を与えた上で高速演算を可能としたセミブラインド MINT 法 [5] も提案されている。

[2] セミブラインド MINT 法 [5] は室内インパルス応答の測定を必要としないが，処理法の前提条件として，話者の方向を正中面にあわせる必要があるため，話者が動く場合は問題になる。

[3] MTF の詳細に関しては，文献 [1] の参考文献 [3]–[6] を参照。

なエンベロープ（振幅あるいはパワー）とキャリア（正弦波あるいは白色雑音）の利用にある[4]。

これらの処理の要点をまとめると，残響音声を回復するために，前者二つの方法 [10, 11] は，残響音声の変調スペクトルの回復を試みているのに対し，後者二つの方法 [12, 13] は残響音声のパワーエンベロープの回復を試みている。これらの方法の特徴は，(1) 室内のインパルス応答を測定することなく残響回復処理を行えること，(2) 音声認識で重要な特徴である振幅情報を回復できること，である。そのため，これらは，応用として非常に有効な手法であると考えられる。しかしながら，彼らは，残響音声のエンベロープ情報（時間的な変動や変調スペクトル）を非常によく回復できたが，結果的に，残響によって低下した音声明瞭度を回復できなかったと報告している [10, 11, 12, 13]。この理由は明白である。彼らは，エンベロープ回復の後，残響の影響を受けたキャリアをそのまま利用して信号を再合成している。再合成された音声は，微細構造で残響の影響を受けたままであり，エンベロープとキャリアの積の不整合が，結果的には人工的な異音を生み出している。そのため，エンベロープ情報は物理的に正しく回復されたが，音声明瞭度は結果的に改善されなかったものと考えられる。

著者の研究グループが取り組んでいる研究の最終的な目標は，伝達系のインパルス応答を測定せずに，ブラインド的に残響音声の物理的な特徴を原音声のもの並に回復し，更に残響によって低下した音声明瞭度も改善可能な回復法を構築することである。第 1 稿 [1] で概説したように，MTF は音声明瞭度に密接な関係があるため，MTF に基づいた時間的なパワーエンベロープ回復法が，明瞭度回復を目指したブラインド的な残響音声回復法を実現するための一つの方法として有効だと考えられる。しかし，上述の問題も含め，パワーエンベロープ逆フィルタ処理をブラインド残響音声回復法まで発展させるためには，まだいくつかの課題が残されている。

次節以降では，MTF に基づくブラインド残響音声回復法を概説し，帯域分割上のパワーエンベロープの回復精度，残響によって低下した音声明瞭度の回復度合，音声認識への効果的な利用を紹介する。

## 3. MTF に基づいたブラインド残響音声回復法

### 3.1 パワーエンベロープ逆フィルタ法

第 1 稿で紹介したパワーエンベロープ逆フィルタ法では，原信号 $\mathbf{x}(t)$，残響信号 $\mathbf{y}(t)$，室内インパルス応答 $\mathbf{h}(t)$ が次のようにモデル化された [1]。

$$\mathbf{x}(t) = e_x(t)\mathbf{n}_x(t) \quad (1)$$

$$\mathbf{h}(t) = e_h(t)\mathbf{n}_h(t) = a\exp(-6.9t/T_R)\mathbf{n}_h(t) \quad (2)$$
$$< \mathbf{n}(t), \mathbf{n}(t-\tau) >= \delta(\tau)$$

$$\mathbf{y}(t) = \mathbf{x}(t) * \mathbf{h}(t) = e_y(t)\mathbf{n}_y(t) \quad (3)$$

[4]広林らは，パワーエンベロープ回復処理が振幅エンベロープ回復処理よりも優れていることを報告している [14]。

ここで，"*" は畳み込み演算，$e_x^2(t)$，$e_h^2(t)$，$e_y^2(t)$ は $\mathbf{x}(t)$，$\mathbf{h}(t)$，$\mathbf{y}(t)$ のパワーエンベロープ，$\mathbf{n}_x(t)$，$\mathbf{n}_h(t)$，$\mathbf{n}_y(t)$ は互いに無相関な白色雑音キャリア（ランダム変数）である。室内インパルス応答のパラーメータ $a$ と $T_R$ は，それぞれ振幅項と残響時間である。この信号モデルの最大の特徴は，各信号の 2 乗集合平均の関係を見たとき，パワーエンベロープ間に畳み込みの関係（$e_y^2(t) = e_x^2(t) * e_h^2(t)$）があることである。ここで，$e_h^2(t)$ の逆フィルタ（IMTF）を利用してパワーエンベロープを回復することから，パワーエンベロープ逆フィルタ処理と呼ばれた（詳細は第 1 稿 [1] を参照）。

### 3.2 帯域分割型パワーエンベロープ逆フィルタ処理

図 1(a) に，文献 [1] で紹介したパワーエンベロープ逆フィルタ法（改良法 [15]）を帯域分割処理に発展させたモデル [16, 17] を示す。まず，原信号 $x(t)$ と残響信号 $y(t)$ が，帯域通過フィルタバンク（$K$ チャネル，$1 \le k \le K$）を利用して分解され，振幅変調形式として

$$x(t) = \sum_{k=1}^{K} x_k(t) = \sum_{k=1}^{K} e_{x,k}(t) \cdot c_{x,k}(t) \quad (4)$$

$$y(t) = \sum_{k=1}^{K} y_k(t) = \sum_{k=1}^{K} e_{y,k}(t) \cdot c_{y,k}(t) \quad (5)$$

と表現されるものと仮定する。但し，$x_k(t)$ と $y_k(t)$ は帯域制限された信号であり，$e_{x,k}(t)$ と $e_{y,k}(t)$，ならびに $c_{x,k}(t)$ と $c_{y,k}(t)$ は，振幅変調の形式を取ったときの，帯域制限された信号のエンベロープとキャリア信号を表す。添字の $k$ はチャンネル番号を示す。

次に，帯域分割におけるパワーエンベロープ逆フィルタ処理を考える。パワーエンベロープ逆フィルタ処理 [15] を利用して，残響音声のパワーエンベロープ $e_{y,k}^2(t)$ から $\hat{e}_{x,k}^2(t)$ に回復する原理の本質は，(1) キャリア $c_{x,k}(t)$ と $e_{y,k}(t)$ が互いに無相関であり，かつ $h(t)$ のキャリア $n_h(t)$ とも無相関であること，(2) $y_k(t) = x_k(t) * h(t)$ の 2 乗集合平均から $e_{y,k}^2(t) = e_{x,k}^2(t) * e_h^2(t)$ が得られることである。既に紹介したように，広林らによって帯域分割型のパワーエンベロープ逆フィルタ法 [13] が考案されているが，文献 [1] で概説したような原理上の問題があるだけでなく，上記のような条件についても深く議論されていない。また，音声の特徴を考慮した帯域分割法についても考察されていない [16]。

本稿では，これらの点について検討した結果の要点を紹介する（詳細については，文献 [15, 16] を参照）

(i) 式 (1) と式 (3) の雑音キャリアが調波複合音であった場合でも，パワーエンベロープ間に畳み込みの関係がほぼ成立する（文献 [15] の 3.3.4 節参照）。

(ii) 帯域制限した音声信号のパワーエンベロープ（$e_{x,k}^2(t)$）間の共変調の関係は，帯域幅を狭めるほど高くなる傾向にある（文献 [16] の 3.1 節参照）。

(iii) $C_{x,k}(t)$ と $C_{y,k}(t)$ を雑音キャリアとしたとき，帯域幅を狭くするほど，かつ $T_R$ が長くなるほど，パワーエンベロープ間の畳み込みの関係が成立しなくなる傾向にある（文献 [16] の 3.2 節参照）。

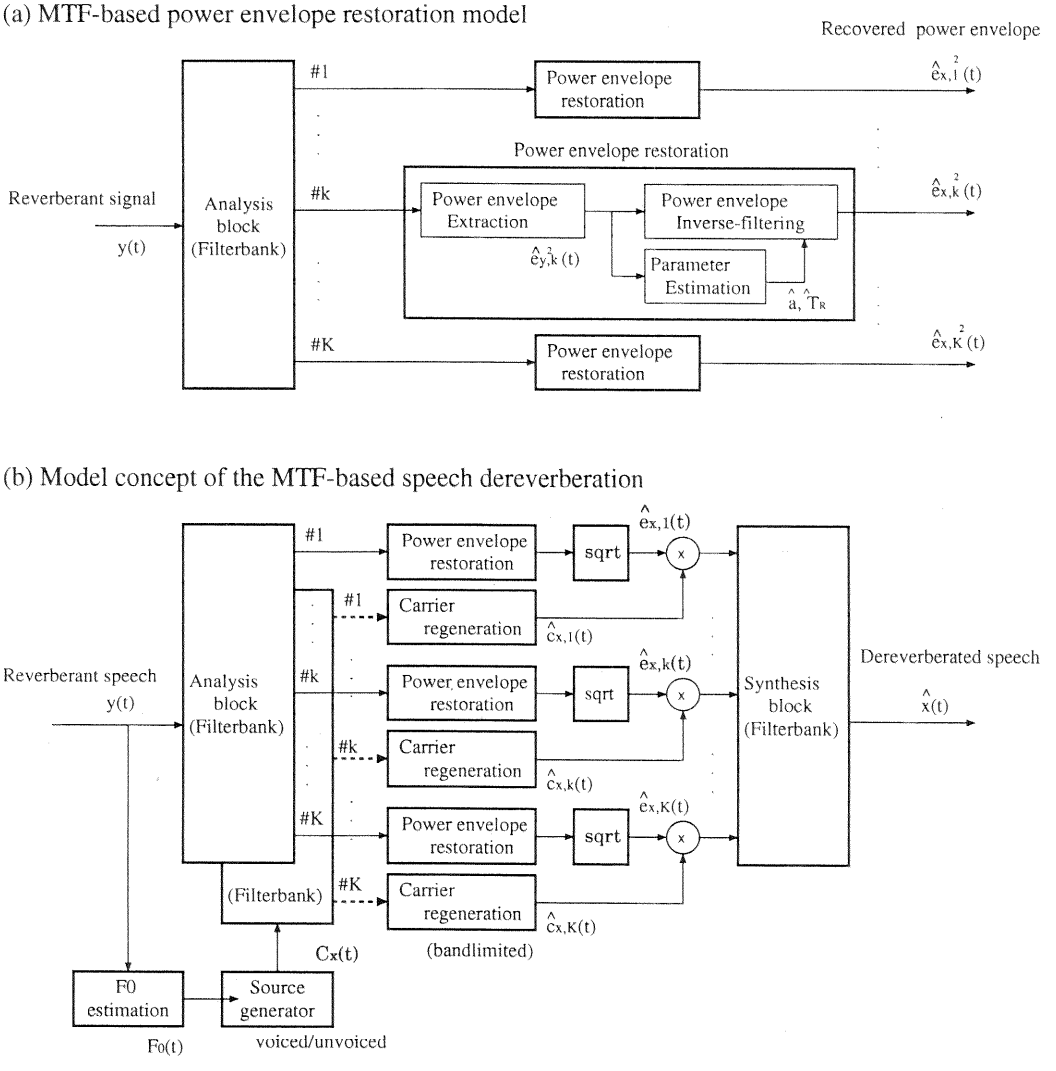



図 1　定帯域幅フィルタバンクにおける (a) パワーエンベロープ逆フィルタ法と (b) ブラインド残響音声回復処理
Fig. 1 (a) Power envelope inverse filtering and (b) blind speech dereveberation method in the constant bandwidth filterbank model

(iv) (ii) と (iii) から，パワーエンベロープ間の共変調と畳み込みの成立にはトレードオフの関係がある。平均的に，100 Hz の帯域分割が最良であった。

(v) 帯域制限した音声信号に対する MTF 概念は，帯域幅が広すぎても狭すぎても成立しない傾向にある（文献 [16] の Fig. 5 参照）。

以上から，帯域分割は，一次近似として一定の帯域幅による分割が好ましく，その帯域幅は 100 Hz 程度が良好であることがわかった[5]。この結果に基づき，本稿では帯域分割型パワーエンベロープ処理を概説する。

### (a) 各チャンネルのパワーエンベロープ $e^2_{y,k}(t)$ の抽出

$$\hat{e}^2_{y,k}(t) = \mathbf{LPF}\left[|y_k(t) + j \cdot \mathbf{Hilbert}(y_k(t))|^2\right] \quad (6)$$

[5] 帯域分割処理に関して，オクターブ帯域幅のような定 Q 帯域幅よりも一定の帯域幅（100 Hz）のほうが適切にパワーエンベロープを回復できることが報告されている [13]。

### (b) 各チャンネルのパワーエンベロープ逆フィルタ法

$$E_{x,k}(z) = \frac{E_{y,k}(z)}{a_k^2}\left\{1 - e^{-\frac{13.8}{\hat{T}_{R,k}\cdot f_s}} z^{-1}\right\} \quad (7)$$

### (c) 各チャンネルの残響時間 $\hat{T}_{R,k}$ の推定

$$\hat{T}_{R,k} = \underset{0\le T_R \le T_{R,\max}}{\arg\min}\left\{\frac{dT_{P,k}(T_R)}{dT_R}\right\} \quad (8)$$

$$T_{P,k}(T_R) = \min\left(\underset{t_{k,\min}\le t\le t_{k,\max}}{\arg\min}\left|\hat{e}_{x,k,T_R}(t)^2 - \theta_k\right|\right).$$

### (d) 各チャンネルの振幅項 $\hat{a}_k$ の決定

$$\hat{a}_k = \sqrt{1/\int_0^T \exp(-13.8t/\hat{T}_{R,k})dt} \quad (9)$$

ここで，**Hilbert()** は Hilbert 変換，$\theta_k$ は $e^2_{y,k}(t)$ の閾値であり，$e^2_{y,k}(t)$ の最大値から −20 dB 低下した値とし

た。$t_{\min}$ と $t_{\max}$ は，それぞれ，$T_{P,k}$ を調べる範囲の下限値と上限値とした。また，低域通過フィルタ $\mathbf{LPF}[\cdot]$ のカットオフ周波数は 20 Hz とした [15]. $T$ は $y(t)$ の信号長，$\hat{e}^2_{x,k,T_R}(t)$ は $T_{R,k}$ の関数として残響回復されたパワーエンベロープの候補の集合，$T_{R,\max}$ は $T_R$ の範囲の上限である。$e^2_{x,k}(t)$，$e^2_h(t)$，$e^2_{y,k}(t)$ の z 変換を，それぞれ，$E_{x,k}(z)$，$E_h(z)$，$E_{y,k}(z)$ である。最終的に $e^2_{x,k}(t)$ は，$E_{x,k}(z)$ の逆 z 変換により得られることになる。ここで，サンプリング周波数 $f_s = 20$ kHz，$K = 100$ の帯域通過フィルタバンクを利用した [16]。各チャンネル毎に $\hat{T}_{R,k}$ と $\hat{a}_k$ を推定しているが，これは全チャンネルで同一の $\hat{T}_R$ と $\hat{a}$ を利用した結果と比較したところ，各チャネル毎に推定したほうが適切であることがわかったためである（詳細は文献 [16] を参照）。

### 3.3 キャリア回復処理への発展

図 1(a) に示した帯域分割型パワーエンベロープ逆フィルタ処理では，信号のパワーエンベロープのみを取り扱っているため，このままでは信号 $\hat{x}(t)$ に復元することができない。先に提案された MTF ベースの残響除去法 [10, 11, 12, 13] では，$y_k(t)$ から $e_{y,k}(t)$ の分解時に得た $c_{y,k}(t)$ を $\hat{x}_k(t)$ の復元に利用していた（$\hat{x}_k(t) = \hat{e}_{x,k}(t) \cdot c_{y,k}(t)$）。そのため，エンベロープとキャリアの不整合から人工的な異音を生み出すこととなり，当初の狙いを達成できなかったものと考えられる。そこで，著者の研究グループでは，キャリア信号 $\hat{c}_{x,k}(t)$ の回復処理を実現することで，不整合を解消し，残響によって低下した音声明瞭度を改善できるものと考えた [17]。現在までに，第 1 稿 [1] で概説した原理から，直接キャリア回復を行う方法が解明されていないが，音声合成技術でよく利用される手法（PIFM 音源モデル [18] や STRAIGHT 音声分析合成器 [19]）を同様に本手法に組み込むことでこの問題を解決することにした。

図 1(b) に，キャリア回復処理を組み込んだブラインド残響音声回復法を示す。これは，パワーエンベロープ回復部とキャリア再生成部（基本周波数推定と音源生成も含む）で構成される。まず，$K$ チャネルのフィルタバンクを通過した残響信号 $y(t)$ をエンベロープ $e_{y,k}(t)$ とキャリア $c_{y,k}(t)$ に分解する。パワーエンベロープ回復部では，前節で説明したパワーエンベロープ逆フィルタ法を用いて，残響音声のパワーエンベロープを回復させる。キャリア再生成部では，残響音声から推定した基本周波数（$F_0$）を基に，有声/無声音ごとに音源信号を再生成し，これがフィルタバンクを通過したものを各チャネルでのキャリアとして再生成する。最後に，各チャネル毎に，回復されたエンベロープ $\hat{e}_{x,k}(t)$ と再生成されたキャリア $\hat{c}_{x,k}(t)$ を掛け合わせ，チャネル信号 $\hat{x}_k(t)$ を作り，合成フィルタバンクによって残響信号から回復した音声 $\hat{x}(t)$ を再構築する。

### 3.4 キャリア再生成処理の実現

図 2 にキャリア再生成処理の流れを示す。周期的な構造を持つ有声音は，$F_0$ と同値の周波数である基本波と $F_0$ の倍数からなる複数の調波を足し合わせた波形構造をもっている。また，基本波は時間的に緩やかに変化し，調波もそれに対応しながら緩やかに変化する。そこで，推定された $F_0$ を利用して調波複合音を作成し，これを有声区間での音源信号のキャリア $\hat{c}_v(t)$ とする。

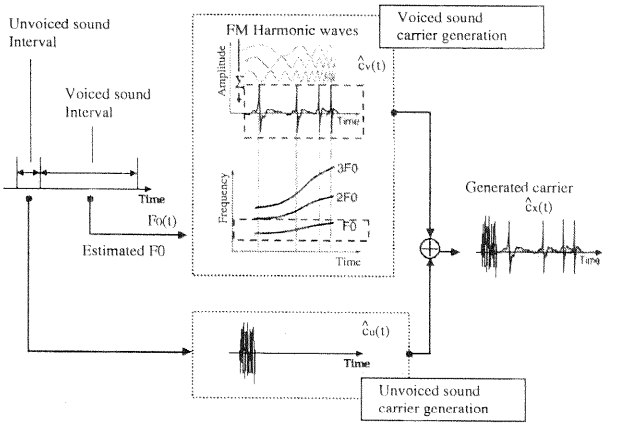



図 2　キャリア再生成処理のブロックダイアグラム
Fig. 2 Blockdiagram of carrier regeneration method

$$\hat{c}_v(t) = \frac{1}{L}\sum_{\ell=1}^{L} \sin\left(\int_0^t 2\pi\ell F_0(\tau)d\tau + \phi_\ell(t)\right) \quad (10)$$

ここで，$F_0(t)$ は基本周波数，$\phi_\ell(t)$ は初期位相，$\ell$ は調波の次数，$L$ は調波の最大次数である。次に，無声音区間のキャリア再生成部では，白色雑音の音源信号であるキャリア $\hat{c}_u(t) = n(t)$ を生成する。これより，全区間の音源信号に対応するキャリア $\hat{c}_x(t)$ を，$\hat{c}_x(t) = \hat{c}_v(t) + \hat{c}_u(t)$ とする。最終的に $\hat{c}_x(t)$ は，分析フィルタバンクによって分解され，各チャネルでのキャリア信号 $\hat{c}_{x,k}(t)$ となる。

提案法では，$F_0(t)$ 推定ならびに有声・無声判断が残響音声 $y(t)$ から正しく行われたものとして，議論を進める。まず，式 (10) の位相項を $\phi_\ell(t) = 0$ として全く制御しないと，合成音特有のバズ音のようなアーティファクトが生じてしまう。これを防ぐために，提案法では，音源信号の位相を直接操作するのではなく，振幅スペクトルとランダム位相スペクトルで構成されるオールパスフィルタを利用して，群遅延を操作することにより，キャリアの位相特性を間接的に操作した。ここでは，STRAIGHT 合成系で利用されたオールパスフィルタ（振幅特性が 1 で，位相特性が次式で表される群遅延 $\tau_g$ を持つフィルタ）を同様に利用した。

$$\tau_g(\omega) = \rho(\omega)\frac{d_g v(\omega)}{\sqrt{\frac{1}{2\pi}\int_{-\pi}^{\pi}|v(\omega)|^2 d\omega}} \quad (11)$$

$$\rho(\omega) = \frac{1}{1+\exp(-(|\omega|-\omega_0)/b_w)} \quad (12)$$

$$v(\omega) = \mathcal{F}^{-1}(W_s(\tau)N(\tau)) \quad (13)$$

$$W_s(\tau) = |\tau|\exp(-\pi(\tau/\tau_{bw})^2) \quad (14)$$

但し，$N(\tau)$ はランダム雑音，$\mathcal{F}^{-1}$ は Fourier 逆変換，$W_s(\tau)$ は荷重関数，$\rho(\omega)$ は周波数 $\omega$ の荷重である。ここで，境界値 $b_w = 3$ kHz と標準偏差 $d_g = 4$ ms の状態で群遅延 $\tau_g$ を設計し，これを有声音のキャリア $c_v(t)$ に畳み込むことで，位相 $\phi_\ell(t)$ を間接的に制御した。

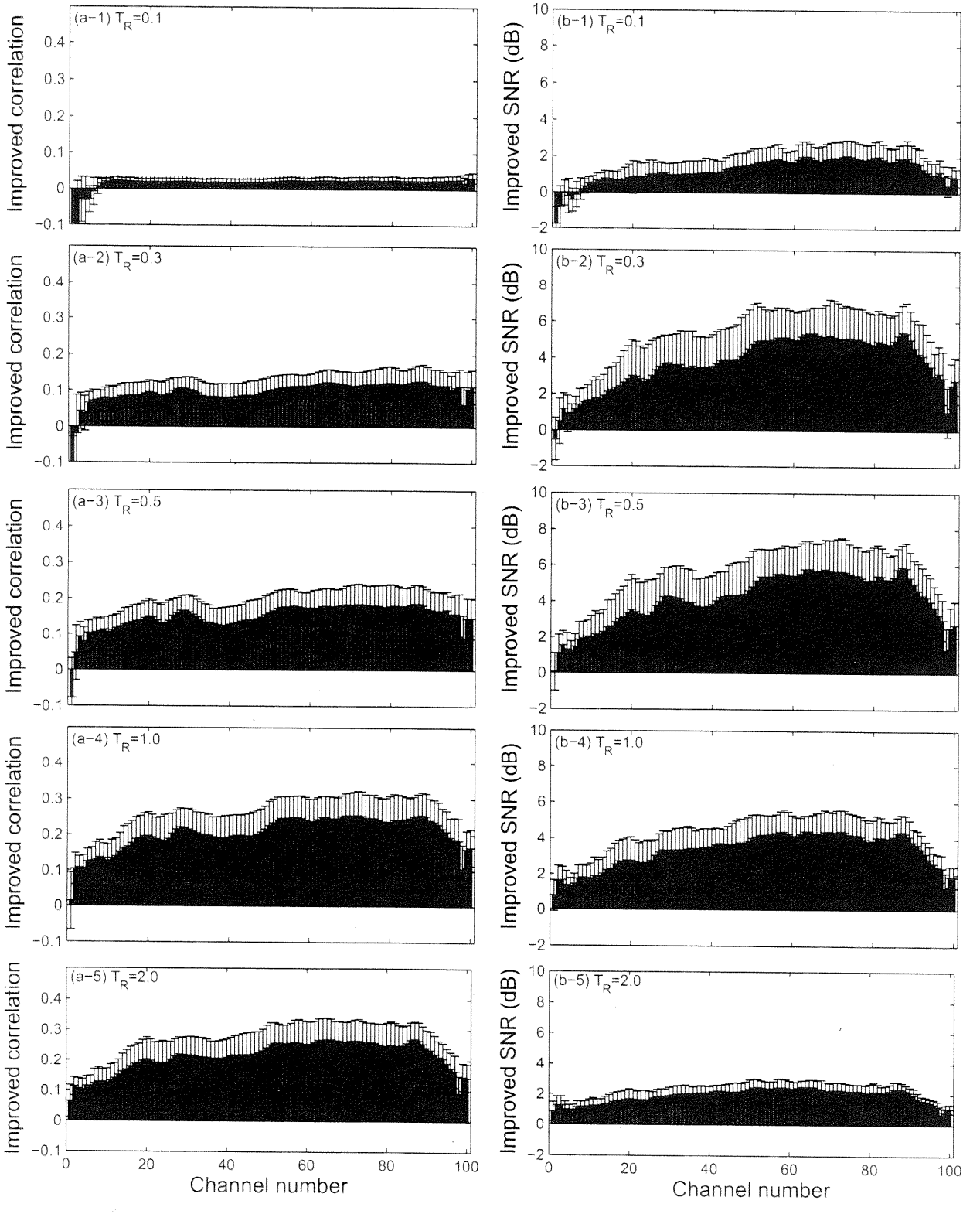



図 3　フィルタバンク処理における音声のパワーエンベロープの回復精度 (a) 相関値の改善度と (b) SNR の改善度（式 8 を利用した場合）

Fig. 3 Improvement in the restoration accuracy for the power envelope of speech on the filterbank: (a) improved correlation and (b) improved SNR (using Eq. (8))

以上が，現在，著者の研究グループで提案されている MTF ベースのブラインド残響音声回復法である。フィルタバンクの構成法（帯域幅の分割法や時間区分化）[20] や変調伝達関数の逆フィルタ（IMTF）の構成法 [21] の検討もあるが，本稿では割愛する。

## 4. 評価シミュレーション

### 4.1 帯域分割型パワーエンベロープ逆フィルタ法の評価

帯域分割型パワーエンベロープ逆フィルタ法を評価するために，ここでは，次のような評価シミュレーションを行う。評価に利用する音声信号は，ATR データベース [22] にある 10 名の話者（男性 5 名（Mau, Mht, Mnm, Mtm, Mtt），女性 5 名（Faf, Ffs, Fkn, Fsu, Fyn））によって発話された日本語単語音声（/aikawarazu/，/shinbun/，/joudan/）とした。室内インパルス応答 $h(t)$ は，式 (2) の人工インパルス応答とし，一つの残響時間 $T_R$ あたり 100 個の白色雑音を用意し，5 つの残響時間（$T_R = 0.1, 0.3, 0.5, 1.0, 2.0$ s）を利用した。そのため，$y(t)$ の総数は，$x(t) * h(t)$ より求めたため，合計 15,000 個（$= 3 \times 10 \times 5 \times 100$）であった。

図 3 は，式 (8) を利用して残響時間 $\hat{T}_{R,k}$ を推定し，残響音声 $y(t)$ の各チャンネルにおけるパワーエンベロープを回復したときの相関値ならびに SNR の改善度を示す（相関値と SNR については，文献 [1] の式 (25) と式 (26) を参照されたい）。$\hat{T}_{R,k}$ はチャンネル毎に別々に推定された。図中では，バーの高さとエラーバーがそれぞれ改善度の平均値と標準偏差を示す。この結果から，$T_R = 0.1$ s を除き，$T_R$ が増加するにつれ，相関値と SNR の改善度が増加していることがわかる。$T_R = 0.1$ s のとき，改善度が若干低くなっている，あるいは負の値になっているのは，残響によるエンベロープの歪みが少なく，それを過剰に回復したためと考えられる。ここでは，図を割愛するが，同様に $\hat{T}_{R,k}$ をチャンネル毎に別々に推定せずに，$\hat{T}_R$ としてすべてのチャンネルで同じ推定値を使った場合，図 3 よりも良い改善値を得ることはなく，逆に改悪されるケースが目立った（詳細については文献 [16] の Fig. 11 を参照）。

図 4 は，図 3 に示した結果の一例を示す。図 4(a) は，男性話者（Mau）によって発話された日本語単語音声 $x(t)$（/aikawarazu/）を示し，図 4(b) は残響音声 $y(t)$（残響時間 $T_R = 1$ s）を示す。図 4(c)-(d) は，このときの各帯域のパワーエンベロープを示すが，図中には全チャンネルではなく，1/4（#1, #5, #9, というように）に間引いて表示している。図中の破線はいずれも原信号のパワーエンベロープ $e^2_{x,k}(t)$ を示している。図 4(c) は帯域分割型パワーエンベロープ逆フィルタ処理を適用しなかった場合の結果（$e^2_{y,k}(t)$，実線）を，図 4(d) は，逆フィルタ処理を適用した場合の結果（$\hat{e}^2_{x,k}(t)$，実線）を示す。図 4(c)-(d) で実線と破線の違いを眺めると，帯域分割型パワーエンベロープ逆フィルタ処理を適用することで，二つの概形がよく一致していることがわかる。

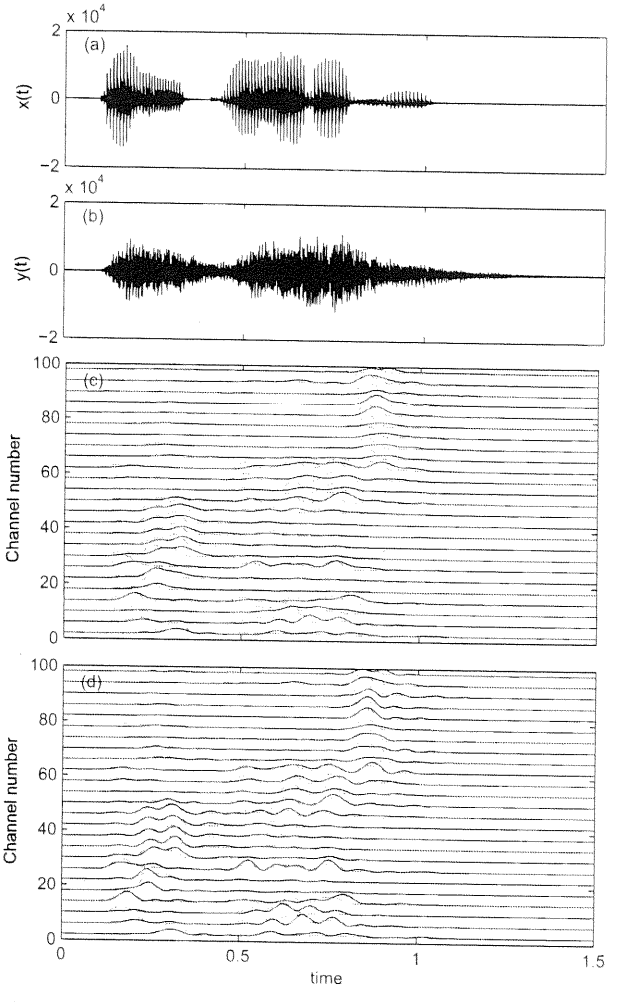



図 4 帯域分割型パワーエンベロープ逆フィルタ処理による残響音声の回復例: (a) $x(t)$ (/aikawaarazu/), (b) $y(t)$ ($T_R = 1.0$ s), (c) $e^2_{y,k}(t)$（実線）と $e^2_{x,k}(t)$（破線），(d) $\hat{e}_{x,k}(t)$（実線）

Fig. 4 Simulation results for the reverberant speech: (a) $x(t)$ (/aikawarazu/), (b) $y(t)$ with $T_R = 1.0$ s, (c) $e^2_{y,k}(t)$ (solid lines) and $e^2_{x,k}(t)$ (dashed lines), and (d) $\hat{e}^2_{x,k}(t)$ (solid lines)

### 4.2 MTF ベースのブラインド残響音声回復処理の評価

図 5 に，図 4 で利用した残響音声（$T_R = 1.0$ s）に対するブラインド残響音声回復の例を示す。図 5(a)-(c) は，それぞれ原音声 $x(t)$，残響音声 $y(t)$，回復音声 $\hat{x}(t)$ の振幅スペクトログラムを示す。図 5(a) と図 5(b) を比較すると，残響は時間–周波数領域における山と谷の間の関係と調波構造を歪ませることがわかる。また，図 5(a) と図 5(c) を比較すると，本手法を利用することで，調波構造と無音区間を効果的に回復できることがわかる。

### 4.3 単語了解度試験による評価

提案法を主観的に評価するために，残響回復音声について音声明瞭度の改善効果を調べた。実験参加者は

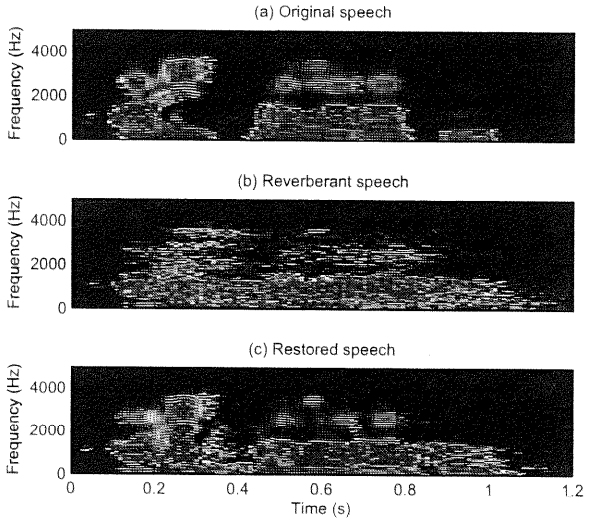



図 5　振幅スペクトログラム: (a) $x(t)$, (b) $y(t)$, (c) 回復音声 $\hat{x}(t)$。$T_R = 1.0$ s.

Fig. 5 Sound spectrogram: (a) $x(t)$, (b) $y(t)$, and (c) restored $\hat{x}(t)$. $T_R = 1.0$ s

正常聴力を有する 5 名の大学院学生であり，防音室内で実験を行った。これらの実験では，親密度別に統制された単語了解度試験用のデータベース [23] 中の 12 単語（各親密度のカテゴリーから 3 単語ずつ）を利用し，各実験刺激は，三つの条件（原音声，残響音声，提案法で回復された音声）とした。これらの刺激に対し，回復音声に対する単語了解度試験を行った。

図 6 に，了解度試験の結果を示す。クリーンな音声のとき，いずれも 100 %の正答率になっているが，残響音声の場合，正答率が低下しているのがわかる。特に，親密度が低下するにつれ，正答率が著しく低下していることがわかる。また，提案法での正答率は，残響音声のものよりも高くなっていることもわかる。このことから，本手法により残響によって低下した音声明瞭度をおおよそ 30%ほど改善できていることがわかる。これは，提案法による回復音声の音質の改善が音声明瞭度の向上に強く寄与しているためと考えられる。

## 5. 応用例：音声認識実験による評価

提案法が残響にロバストな音声認識の前処理として効果的に働くことを確認するために，次の音声認識実験を行った。音声データとして，AURORA-2J データベース [24] を使った。このうち，8840 個のクリーンな音声を音響モデルの学習のために利用し，残りの 1001 個に室内インパルス応答 $h(t)$ を畳み込んで得られた残響音声をテスト用に利用した。テストでは 2 種類の室内インパルス応答を利用した。一つは，人工残響環境として，式 (2) のインパルス応答を利用し，残響時間 $T_R$ を 0.0〜2.0 s まで，0.2 s 刻で制御した合計 21 個を用意した。もう一つは，SMILE2004 で提供している実環境の室内インパルス応答 [25]（43 個）を利用した。

この実験では，データベースのサンプリング周波数 $f_s = 8$ kHz にあわせるため，帯域分割数は，帯域幅を

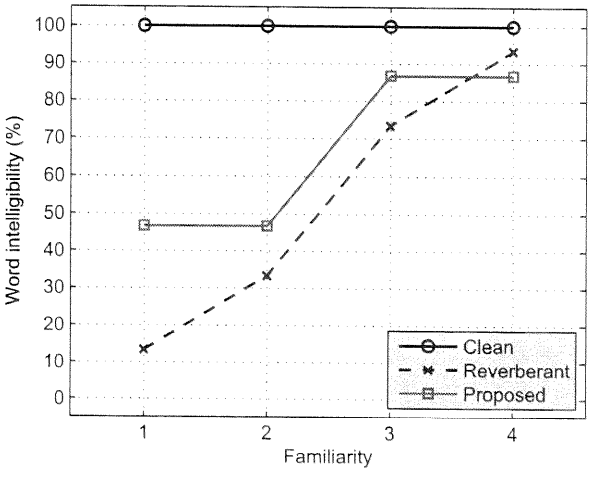



図 6　単語了解度試験の結果

Fig. 6 Results of word intelligibility tests

100 Hz 固定として，$K = 40$ とした。ここで，認識に利用する特徴は，図 1(a) の出力であるパワーエンベロープ $\hat{e}^2_{x,k}(t)$ に対するメル周波数ケプストラム係数（MFCC）である。これを得る手順は次のとおりである。はじめに，$e^2_{x,k}(t)$ に対し，次式に示す平滑化処理を適用した。

$$\bar{e}_{x,k}[n] = \lambda \bar{e}_{x,k}[n-1] + (1-\lambda)\hat{e}_{x,k}[n] \qquad (15)$$

この平滑化は良く知られた漏洩積分器の一つであり，ここでは $\lambda = 0.98$ とした。次に，平滑化されたエンベロープに対し，フレーム長 32 ms，フレームシフト 16 ms，ハミング窓を利用したフレーム積分処理を適用して，スペクトルに相当するエンベロープを得た。最後に，このスペクトルに対し圧縮処理を行い，全チャンネルに渡って離散コサイン変換（DCT）することで MFCC を求めた。認識に使う特徴ベクトルは 39 次元で構成され，最初の 13 次は 0 次の対数パワー項を含む 12 次の MFCC である。残りの 26 次は，このベクトルのそれぞれ 1 階差分と 2 階差分（$\Delta$ と $\Delta\Delta$）である。

音響モデルは，10 個の数字に対する HMM，一つの無音に対する HMM，小休止に対する HMM で構成される。これらは，AURORA-2J 実験で利用されたものと全く同じものである [24]。各数字に対する HMM は，16 分布，18 状態をもつ。無音に対する HMM は 3 分布，5 状態をもち，小休止に対する HMM は 1 分布，3 状態をもつ。数値の各分布は，20 次のガウス混合から成るのに対し，無音と小休止では 36 次のガウス分布で構成された。ここでは，HMM 音響モデルを構築するために，HTK 音声ツール [26] を利用した。

はじめに人工残響環境における認識実験の結果を図 7 に示す。比較のため，同じ条件下で二つの簡便法（ケプストラム平均による正規化法（CMN）[27] と RASTA フィルタ法 [28]）を利用したときの結果を示す。図中の CBFB は，提案法で利用している定帯域幅フィルタバンクを示し，CQFB は定 Q フィルタバンクで同様の処理をした場合を示す（詳細は文献 [29] を参照）。

図 7(a) からわかるように，残響時間 $T_R$ が長くなるにつれ，前処理なしの認識率（CBFB, CQFB）が著しく

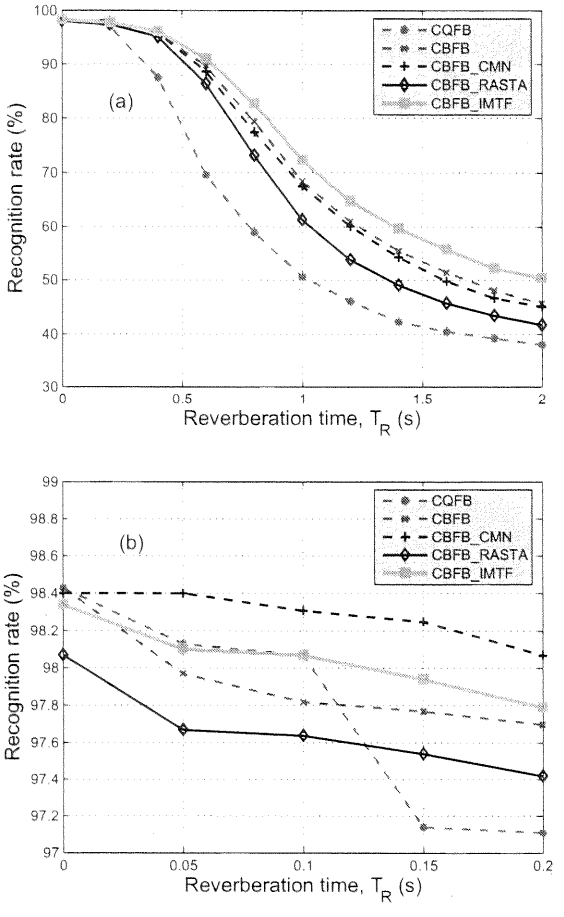



図 7　人工的な残響環境における残響音声の認識率の比較評価 (a) 全体評価と (b) $T_R = 0.0$〜$0.2$ s にズームアップした残響時間内での評価

Fig. 7 Comparative evaluations of reverberant speech recognition rates: (a) whole evaluation and (b) close-up of plot in range from 0.0 to 0.2 s

低下することがわかる。また，CMN ならびに RASTA 法を利用したものも同様に低下していることがわかる。図 7(b) は，残響時間が短い場合 ($T_R < 0.2$ s) の結果を示す。ここでは，CMN（CFBF_CMN）が効果的に働いているが，RASTA 法（CBFB_RASTA）は効果的に働いていないことがわかる。これに対し，提案法では，残響時間が長くなるにつれ，MTF に基づくエンベロープ回復（CBFB_IMTF）が認識率の低下を防いでいることがわかる。ここで，相対的な認識率の改善度（RI）として，次式を利用する [29]。

$$\mathrm{RI} = \frac{(\mathrm{TRR} - \mathrm{BRR})}{(1 - \mathrm{BRR})} \times 100 \quad (\%) \qquad (16)$$

但し，"TRR" と "BRR" は，テスト時の認識率とベースラインでの認識率である。ここでは，ベースラインとして CQFB の結果を利用した。図 7 の結果では，提案法（CBFB_IMTF）が，$0.2 < T_R < 2$ s の範囲で，28.64%〜35.67%の改善度を示した。

最後に，表 1 に実際の残響環境における認識結果を示す。テスト時の残響時間は，短いもので 0.3 s 程度，長いもので 3 s 程度であった。比較結果の中で最良な認識結果を太字で示す。従来の MFCC を利用した場合，ならびに前処理なしの比較法（CBFB と CQFB）を利用した場合，いずれも類似した値であり，比較対象の中で悪い結果となった。これに対し，従来法（CMN と RASTA 法）を利用した場合，先の結果より改善される結果となったが，それほど大きな改善度にはなっていない。これらに対し，提案法の結果（CBFB_IMTF）は，ほとんどの条件で最良な結果を示している。最良な結果になっていない場合は，残響時間が短い場合であり，これらの認識率よりも若干低い程度であった。

## 6. 残された課題

前節にて，MTF ベースのブラインド残響音声回復法の処理効果として，残響によって低下した音声明瞭度（単語了解度）ならびに音声認識率の改善効果を示した。いずれも残響による影響が取り除かれ，ある程度の改善効果が見られたが，絶対的な改善レベルまでには到達しなかった。例えば，単語了解度試験の結果では，親密度が低い場合で了解度を 100%までに回復できなかったし，音声認識実験では，残響時間が 1 s を越える状態で，音声認識を現実的に利用できる程度（認識率が 90%を超える）までには至らなかった。

今後，これらを改善するためには，次の課題を検討しなければならい：(1) パワーエンベロープ逆フィルタ法に最も適切なフィルタバンクの構築，(2) 残響時間推定の不一致とそれに伴うパワーエンベロープ回復の低減の改善，(3) 時間–周波数区分化を考慮したパワーエンベロープ回復，(4) 残響にロバストな基本周波数推定法ならびに有声・無声判定，(5) 音声合成器の音質改善。

## 7. おわりに

本稿では，変調伝達関数（MTF）に基づく音声信号処理（全 3 回シリーズ）の第 2 稿として，MTF ベースのブラインド残響音声回復法を概説し，現状でどの程度のことまで達成されたのか，また残された課題が何であるかを説明した。次回（第 3 稿）は，本稿で紹介したブラインド残響音声回復法の問題点を解決するための方向で派生した研究課題として，残響環境下での基本周波数推定法と残響時間推定法を概説する。

## 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。

表 1　実際の残響環境 [25] における残響音声の認識率 (%)：残響時間 $T_R$ は 125 Hz から 8 kHz のオクターブ周波数内での伝達関数のすべての残響時間の平均値として求められた。"RI_CQFB" と "RI_CBFB" は CQFB と CBFB の特徴での認識率と比較したときの相対的なエラー減少率を示す。

Table 1 Reverberant speech recognition rates (%) in actual reverberant environments[25]: The reverberation time, $T_R$, was determined as the average of all $T_R$s on the transfer function at 125 Hz to 8 kHz in octave frequencies. "RI_CQFB" and "RI_CBFB" mean the relative improvement in the error reduction rate of the CBFB_IMTF feature compared with those of CQFB and CBFB features

| 室の条件 (RIR) | 残響時間 $T_R$ (s) | MFCC | CQFB | CBFB | CBFB_ CMN | CBFB_ RASTA | CBFB_ IMTF | RI_ CQFB | RI_ CBFB |
|---|---|---|---|---|---|---|---|---|---|
| 多目的ホール 1 [1] | 1.09 | 42.55 | 45.56 | 52.44 | 57.63 | 48.51 | **60.30** | *27.08* | *16.53* |
| 多目的ホール 1 [2] | 0.80 | 55.39 | 54.31 | 68.52 | 71.85 | 66.17 | **74.33** | *43.82* | *18.46* |
| 多目的ホール 2 [3] | 1.44 | 32.88 | 36.60 | 40.62 | 39.70 | 32.64 | **45.41** | *13.90* | *8.07* |
| 多目的ホール 2 [4] | 1.04 | 39.70 | 43.51 | 47.56 | 45.49 | 36.20 | **52.38** | *15.70* | *9.19* |
| 多目的ホール 3 [5] | 1.93 | 30.70 | 33.40 | 33.80 | 35.31 | 31.26 | **39.31** | *8.87* | *8.32* |
| 多目的ホール 3 [6] | 1.35 | 42.12 | 43.48 | 46.52 | 53.42 | 47.50 | **54.19** | *18.95* | *14.34* |
| 多目的ホール 4 [7] | 1.42 | 55.70 | 55.07 | 69.63 | 74.24 | 71.05 | **75.87** | *46.29* | *20.55* |
| 多目的ホール 4 [8] | 1.54 | 52.44 | 53.42 | 67.02 | 71.08 | 66.78 | **73.10** | *42.25* | *18.44* |
| 多目的ホール 5 [9] | 1.47 | 46.55 | 47.28 | 61.38 | 59.84 | 54.71 | **64.04** | *31.79* | *6.89* |
| 多目的ホール 6 [10] | 2.16 | 40.13 | 42.83 | 49.95 | 49.43 | 47.99 | **54.49** | *20.40* | *9.07* |
| コンサートホール 1 [11] | 2.35 | 27.72 | 34.20 | 35.19 | 33.50 | 28.92 | **35.92** | *2.61* | *1.13* |
| コンサートホール 1 [12] | 2.34 | 30.09 | 35.65 | 39.88 | 37.03 | 33.22 | **42.74** | *11.02* | *4.76* |
| コンサートホール 1 [13] | 2.35 | 30.40 | 35.22 | 37.67 | 35.34 | 33.19 | **43.17** | *12.27* | *8.82* |
| コンサートホール 1 [14] | 2.39 | 30.58 | 35.37 | 39.73 | 38.44 | 35.55 | **45.47** | *15.63* | *9.52* |
| コンサートホール 1 [15] | 2.38 | 27.82 | 33.93 | 36.17 | 34.30 | 32.36 | **40.56** | *10.03* | *6.88* |
| コンサートホール 2 [16] | 1.14 | 40.34 | 44.34 | 50.60 | 58.12 | 49.59 | **59.84** | *27.85* | *18.17* |
| コンサートホール 3 [17] | 1.96 | 35.00 | 36.81 | 37.73 | 42.80 | 39.12 | **46.33** | *15.07* | *13.81* |
| コンサートホール 4 [18] | 1.92 | 41.23 | 41.42 | 50.02 | 49.95 | 46.15 | **54.38** | *22.12* | *8.72* |
| コンサートホール 4 [19] | 2.55 | 34.33 | 36.72 | 41.97 | 41.14 | 37.15 | **44.43** | *12.18* | *4.24* |
| コンサートホール 5 [20] | 2.32 | 31.78 | 37.70 | 38.29 | 34.85 | 32.58 | **44.09** | *10.19* | *9.40* |
| コンサートホール 6 [21] | 1.77 | 37.73 | 41.42 | 43.57 | 42.55 | 38.38 | **53.45** | *20.54* | *17.51* |
| コンサートホール 6 [22] | 1.74 | 40.13 | 44.18 | 47.87 | 46.27 | 42.25 | **55.14** | *19.63* | *13.95* |
| コンサートホール 6 [23] | 1.69 | 34.73 | 38.23 | 44.34 | 43.11 | 41.42 | **52.69** | *23.41* | *15.00* |
| 教室 [24] | 1.36 | 46.76 | 45.72 | 60.85 | **70.31** | 67.58 | 68.53 | *42.02* | *19.62* |
| 劇場 [25] | 0.85 | 46.24 | 48.82 | 60.55 | 60.39 | 53.39 | **63.68** | *29.03* | *7.93* |
| 会議室 [26] | 0.62 | 77.43 | 72.24 | 89.10 | 91.25 | 89.16 | **91.62** | *69.81* | *25.87* |
| 教室 [27] | 1.12 | 55.85 | 53.18 | 70.83 | **81.12** | 78.75 | 80.32 | *57.97* | *32.53* |
| 教室 [28] | 1.09 | 57.48 | 51.30 | 68.35 | **83.97** | 80.75 | 78.85 | *56.57* | *33.18* |
| スピーチホール 1 [29] | 1.54 | 40.44 | 44.89 | 51.58 | 46.58 | 44.55 | **54.34** | *17.15* | *5.70* |
| 教会 1 [30] | 0.71 | 57.35 | 56.95 | 70.34 | 76.60 | 72.43 | **77.56** | *47.87* | *24.34* |
| 教会 2 [31] | 1.30 | 33.71 | 37.21 | 41.42 | 40.87 | 30.52 | **42.49** | *8.41* | *1.83* |
| イベントホール 1 [32] | 3.03 | 27.51 | 31.19 | 33.40 | 33.40 | 30.80 | **36.87** | *8.25* | *5.21* |
| イベントホール 2 [33] | 3.62 | 28.77 | 32.98 | 35.62 | 37.27 | 34.88 | **41.63** | *12.91* | *9.34* |
| 体育館 1 [34] | 2.82 | 21.61 | 26.59 | 29.08 | 27.88 | 25.39 | **30.09** | *4.77* | *1.42* |
| 体育館 2 [35] | 1.70 | 32.51 | 37.33 | 39.98 | 41.60 | 36.29 | **48.23** | *17.39* | *13.90* |
| リビングルーム [36] | 0.36 | 89.81 | 86.40 | **98.31** | 96.75 | 95.30 | 96.90 | *77.21* | *-83.93* |
| 映画館 [37] | 0.38 | 88.36 | 84.22 | 93.49 | **95.95** | 92.85 | 93.18 | *56.78* | *-4.76* |
| アントリウム [38] | 1.57 | 35.19 | 36.91 | 39.70 | 43.97 | 36.08 | **48.60** | *18.53* | *14.76* |
| トンネル [39] | 2.72 | 28.52 | 25.05 | 25.33 | 26.76 | **35.06** | 33.87 | *11.77* | *11.44* |
| 駅のコンコース [40] | 1.95 | 36.66 | 39.64 | 44.06 | **46.18** | 34.48 | 45.93 | *10.42* | *3.34* |
| スピーチホール 2 [41] | 1.53 | 38.26 | 41.45 | 48.33 | 46.88 | 42.80 | **56.13** | *25.07* | *21.10* |
| スピーチホール 2 [42] | 1.49 | 34.26 | 37.67 | 45.13 | 44.98 | 41.26 | **51.77** | *22.62* | *12.10* |
| スピーチホール 2 [43] | 1.40 | 39.73 | 39.05 | 54.41 | 59.81 | 56.19 | **65.18** | *42.87* | *23.62* |

[1] 反射板あり，体積 2,000 $m^3$，[2] 反射板なし，[3] 反射板あり，体積 5,700 $m^3$，[4] 反射板なし，[5] 反射板あり，体積 7,200 $m^3$，[6] 反射板なし，[7] 吸収板あり，体積 12,000 $m^3$，[8] 吸収板なし，[9] 体積 14,000 $m^3$，[10] 体積 19,000 $m^3$，[11] 体積 5,600 $m^3$，[12] スピーカー・マイクロホン間距離 $d = 6$ m，[13] $d = 11$ m，[14] $d = 15$ m，[15] $d = 19$ m，[16] 体積 6,100 $m^3$，[17] 体積 20,000 $m^3$，[18] 吸収カーテンあり，体積 7,100 $m^3$，[19] 吸収カーテンなし，[20] 体積 17,000 $m^3$，[21] 1F 正面，体積 17,000 $m^3$，[22] 2F，[23] 3F，[24] フラッターエコーあり，[25] 体積 3,900 $m^3$，[26] 体積 130 $m^3$，[27] 体積 400 $m^3$，[28] 体積 2,400 $m^3$，[29] 体積 11,000 $m^3$，[30] 体積 1,200 $m^3$，[31] 体積 3,200 $m^3$，[32] 体積 28,000 $m^3$，[33] 体積 41,000 $m^3$，[34] 体積 12,000 $m^3$，[35] 体積 29,000 $m^3$，[36] 木製，体積 110 $m^3$，[37] 体積 560 $m^3$，[38] 体積 4,000 $m^3$，[39] 体積 5,900 $m^3$ 距離 120 m，[40] 駅構内，[41] 1F 正面，[42] 1F 中央，[43] バルコニー

## 参考文献


[1] 鵜木祐史: 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—, 信号処理学会誌, Vol. 12, No. 5, pp. 339-348, 2008.

[2] S. T. Neely and J. B. Allen: Invertibility of a room impulse response, J. Acoust. Soc. Am., Vol. 66, pp. 165–169, 1979.

[3] M. Miyoshi and Y. Kaneda: Inverse filtering of room acoustics,IEEE Trans. ASSP, Vol. 36, pp. 145–152, 1988.

[4] H. Nakajima, M. Miyoshi and M. Tohyama: Sound field control by indefinite MINT filter, IEICE Trans. Fundamentals, Vol. E80-A, No. 5, pp. 821–824, 1997.

[5] 古家賢一, 片岡章俊: チャネル間相関行列と音声の白色化フィルタを用いた Semi-blind 残響抑圧, 信学論 A, Vol. J88-A, No. 10, pp. 1089–1099, 2005.

[6] H. Wang and F. Itakura: Realization of acoustic inverse filtering through multi-microphone sub-band processing, IEICE Trans. Fundamentals, Vol. E75-A, pp. 1474–1483, 1992.

[7] T. Nakatani and M. Miyoshi: Blind dereverberation of single channel speech signal based on harmonic structure, Proc. ICASSP'03, Vol. 1, pp. 92–95, 2003.

[8] T. Nakatani, K. Kinoshita and M. Miyoshi: Harmonicity-Based Blind Dereverberation for Single-Channel Speech Signals, IEEE Trans. Audio. Speech, and Language Processing, Vol. 15, No. 1, pp. 80–95, 2007.

[9] K. Kinoshita, T. Nakatani and M. Miyoshi: Harmonicity Based Dereverberation for Improving Automatic Speech Recognition Performance and Speech Intelligibility, IEICE Trans. Fundamentals, Vol. E88-A, No. 7, pp. 1724–1731, 2005.

[10] T. Langhans and H. W. Strube: Speech enhancement by nonlinear multiband envelope filtering, Proc. ICASSP82, pp. 156–159, 1982.

[11] C. Avendano and H. Hermansky: Study on the dereverberation of speech based on temporal envelope filtering, Proc. ICSLP96, pp. 889–892, 1996.

[12] J. Mourjopoulos and J. K. Hammond: Modelling and enhancement of reverberant speech using an envelope convolution method, Proc. ICASSP83, pp. 1144–1147, 1983.

[13] 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, 信学論 A, Vol. J83-A, No. 8, pp. 1029–1033, 2000.

[14] 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.

[15] 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.

[16] 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.

[17] M. Unoki, K. Sakata and M. Akagi: A speech dereverberation method based on the MTF concept, Proc. EuroSpeech2003, pp. 1417–1420, 2003.

[18] 濱上 知樹: 音源波形形状を高調波位相により制御する音声合成方式, 日本音響学会誌, Vol. 54 No. 9, pp. 623-632, 1998.

[19] H. Kawahara, I. Masuda-Katsuse and A. de Cheveign'e: Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction, Speech Communication, Vol. 27, pp. 187–207, 1999.

[20] M. Unoki, M. Toi and M. Akagi: Development of the MTF-based speech dereverberation method using adaptive time-frequency division, Proc. Forum Acusticum 2005, pp. 51–56, Budapest, Hungary, 2005.

[21] M. Unoki, M. Toi and M. Akagi: A speech dereverberation method based on the MTF concept using adaptive time-frequency divisions, *Proc. SPECOM2006*, pp. 1689–1692, St. Petersburg, Russia, 2006.

[22] T. Takeda, Y. Sagisaka, K. Katagiri, M. Abe and H. Kuwabara: Speech Database User's Manual, ATR Technical Report, TR-I-0028, 1988.

[23] Database for speech intelligibility testing using Japanese word lists. NTT-AT, 2003.

[24] AURORA-2J: http://sp.shinshu-u.ac.jp/CENSREC/.

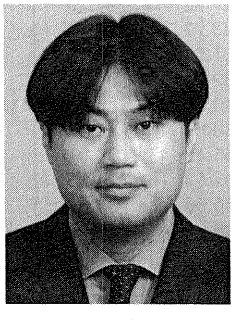
[25] SMILE2004, Sound Material in Living Environment, Architectural Institute of Japan and GIHODO SHUPPAN Co., Ltd., 2004.

[26] The HTK Book (version 3.2), Cambridge University Engineering Department, 2002.

[27] F. Liu, R. Stern, X. Huang and A. Acero: Efficient cepstral normalization for robust speech recognition, Proc. ARPA Human Language Technology Workshop, 1993.

[28] H. Hermansky, N. Morgan and H. G. Hirsch: Recognition of speech in additive and convolutional noise based on RASTA spectral processing, Proc. ICASSP'93, pp. 83–86, 1993.

[29] X. Lu, 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, Acout. Sci. & Tech., Vol. 29, No. 6, pp. 351–361, 2008.


**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。