変調伝達関数に基づく音声信号処理 (3)
—残響環境下の基本周波数推定法と残響時間のブラインド推定—
Speech Signal Processing Based on the Concept of Modulation Transfer Function (3)
— Robust F_0 Estimation Method and Blind Estimation of Reverberation Time in Reverberant Environments—

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。次に，第２稿 [2] 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た。本稿（第３稿）では，第２稿で紹介した研究課題から派生した二つの課題として，MTF に基づいた基本周波数推定法と残響時間推定法を概説し，どの程度のことが現在までに達成されているのか，何が残された課題であるかを説明する。

2. 基本周波数推定法の取組み

音声の基本周波数（F_0）[1]は，音声の音源情報（声帯振動）を表すため，音声分析合成，音声認識，音声強調，声質変換といった様々な音声信号処理において重要な特徴として利用されている。そのため，実環境においてロバストで正確な F_0 推定法が必要とされている。

F_0 推定法については，約半世紀近くに渡って，多くの検討がなされ，種々の手法が提案されてきた（例えば，文献 [3, 4]）[2]。これらの推定法は，大雑把に，時間領域では音声信号の周期性の特徴（例えば，零交差，ピーク検出法，自己相関法，AMDF 法），周波数領域では調波性の特徴（例えば，comb フィルタリング法，自己相関法，ケプストラム法）を利用するものに分類される [3, 4, 5]。これらの主な狙いは，観測した音声信号から音源情報の周期性・調波性の特徴を取り出すことである。しかし，次にあげる三つの問題点のため，この狙いは完全には達成されていない。

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) **観測可能性**：観測された音声信号は口唇・鼻腔から放射されるため，声道特性を取り除かない限り，観測された音声信号から音源情報のみを正確に抽出できないこと。

(2) **変動性と不規則性**：声帯振動は完全な周期信号ではなく，その変動範囲がかなり広いこと。

(3) **ロバスト性**：観測された音声信号は雑音や残響の影響を受けるため，F_0 推定に必要な情報が汚され，その推定精度を低下させる。そのため，環境等の影響を受けず頑健に F_0 情報を抽出できなければいけないこと。

これまでの多くの検討では，暗黙の了解で，音声信号は雑音がない状態か，クリーンな環境で観測されるものとしていたため，最初の二つの問題点に焦点が集まっていた。そのため，まずは音源フィルタモデルに基づき，観測された音声信号からフィルタ特性を取り除くことで，問題点 (1) を解決する手法が提案されてきた（例えば，準同形処理法 [6] や LPC 法 [7]）。最近では，音声信号のロバストな特徴に着目し，瞬時周波数の不動点を利用する方法（TEMPO [8]）や自己相関法と AMDF 法の組合せによる改良法（YIN [9]）が知られている。両方法とも，極めて正確に目的音声の F_0 を推定できることから，最初の二つの問題点は概ね解決されたものと考えられる。しかし，これらの方法が実環境下で正確に F_0 を推定できるかわかっていない。そのため，問題点 (3) を早急に検討する必要がある。

現在までに，音声信号の周期性・調波性を利用した F_0 推定法が雑音環境下でロバストであることが報告されている [4, 10, 11]。音声信号の瞬時振幅は，音源情報の周期性・調波性の特徴を有している。また音声信号の瞬時周波数は，正確な F_0 推定に有効なパラメータとして利用されている。しかし，TEMPO で利用されて

[1] 基本周波数をピッチと呼ぶケースがあるが，正確には，基本周波数は物理用語，ピッチは “音の高さ” を指す心理用語である。基本周波数は音の高さに密接な関係があるため，このような混同された使い方がなされてきたと考えられるが，本稿では啓蒙活動の一つとして，ピッチとは呼ばず，明確に “基本周波数” と呼ぶことにする。

[2] F_0 推定を広くサーベイした論文あるいは書籍は多くない。Hess による書籍 [3] あるいはチャプター論文 [4] がそれにあたる。

いる瞬時周波数に基づく不動点は雑音に敏感であるため，雑音環境下においては，TEMPO の耐雑音性の低下を招いている。これを改良する方法として，調波性と帯域幅方程式を利用した方法 [10] や瞬時振幅・周波数に関係した周期性・調波性を利用した方法 [11] が提案された。音声信号の周期性・調波性を利用したこれらの F_0 推定法が，雑音環境下で TEMPO よりもロバストで正確に F_0 を推定できることが報告されている。

これらすべての方法は，目的音声の F_0 を十分正確に推定するために，クリーンな環境か雑音環境でのみ検討されてきた。そのため，瞬時振幅と瞬時周波数，あるいは，雑音に頑健な特徴である周期性や調波性を利用することで，雑音音声から F_0 を正確に，頑健に推定できるようになった。この点で，問題点 (3) の一部（雑音環境に関すること）を解決できたものと考えられるが，これらの方法が実際の残響環境でも十分に機能するかはわかっていない。

著者の研究グループでは，上記の疑問に答えるために，大規模な音声データセットに対し，代表的な基本周波数推定法が人工的な残響環境ならびに実際の残響環境においてどれだけ正確に基本周波数を推定できるか検討した [12, 13]。その結果，評価で利用した代表的な方法すべてが，残響環境でほとんど機能せず，特に残響時間が長くなると推定精度が著しく低下することを明らかにした。これらの評価・分析を基に，著者の研究グループでは，複素ケプストラム分析において，音源フィルタモデルと変調伝達関数の概念を利用することで，残響音声から F_0 を推定する方法を提案した [12, 13]。

本稿では，残響環境下のみに限定するが，大規模な音声データと実残響インパルス応答データを利用して，著者の研究グループで既に提案された方法ならびに代表的な F_0 推定法の残響環境下での推定精度を比較評価し，提案法の有効性ならびに残された課題を説明する。本提案法は，例えば，第 2 稿 [2] の第 3.3 節で説明したような MTF ベースの残響音声回復法におけるキャリア再生成で重要な役割を果たすことになる。

3. MTF に基づいた基本周波数推定法

3.1 基本周波数推定の問題設定

時変な調波信号 $x(t)$ が次式で表されるものとする。

$$x(t) = \sum_{k \in K} a_k(t) \exp(j\omega_k(t)t + \theta_k(t)) \qquad (1)$$

ここで，$a_k(t)$ は瞬時振幅，$\theta_k(t)$ は瞬時位相，k は調波の次数，K はその最大次数である。また，角周波数は $\omega_k(t) = 2\pi k F_0(t)$ であることから，基本周波数 $F_0(t)$ は瞬時周波数として表される。これに対し，残響環境下では，$x(t)$ からではなく，次式で表される残響音声 $y(t)$ あるいはその短時間 Fourier 変換（STFT）$Y(\omega,\tau)$ から $F_0(t)$ を推定することが主目的となる。

$$\begin{aligned} y(t) &= x(t) * h(t) = e(t) * v_\tau(t) * h(t) \qquad (2) \\ Y(\omega,\tau) &= X(\omega,\tau)H(\omega,\tau) \\ &= S(\omega,\tau)V(\omega,\tau)H(\omega,\tau) \qquad (3) \end{aligned}$$

但し，$X(\omega,\tau)$ と $H(\omega,\tau)$ は，それぞれ，調波信号 $x(t)$ と室内インパルス応答（残響特性）$h(t)$ の STFT である。$e(t)$ と $v_\tau(t)$ は，それぞれ，音源フィルタモデルにおける音源信号（声帯振動）と，時刻 τ でのフィルタ特性（声道特性）のインパルス応答である。また，$S(\omega,\tau)$ と $V(\omega,\tau)$ は，それぞれ，$e(t)$ と $v(t,\tau) = v_\tau(t)$ の STFT である。ここで，$H(\omega,\tau)$ は，長時間 Fourier 変換 (LTFT) から得られたフィルタ特性（$H(\omega) = H(\omega,\tau)$）をすべて含まなければならないため，分析幅は残響時間よりも十分に長いものとする。

3.2 複素ケプストラム分析

式 (3) から，残響信号 $y(t)$ の複素ケプストラムは

$$C_Y(q,\tau) = C_{\mathrm{src}}(q,\tau) + C_{\mathrm{flt}}(q,\tau) + C_H(q,\tau) \qquad (4)$$

と表される。ここで，$C_H(q,\tau)$ は残響インパルス応答 $h(t)$ の複素ケプストラムである。$C_{\mathrm{src}}(q,\tau)$ と $C_{\mathrm{flt}}(q,\tau)$ は，それぞれ，音源特性とフィルタ特性の複素ケプストラムである。これらは，振幅ケプストラムと位相ケプストラムに分離して表すこと（添字に "A"，"ϕ" と記述）もできる。更に，これらは，最小位相成分と非最小位相成分（all-pass 位相成分）に分離して表すこと（添字に "min"，"all" と記述）も可能である。更に，$|X_{\mathrm{all}}(\omega,\tau)| = 1$ と $C_{A,\mathrm{all}}(q,\tau) = 0$ であることから，$C_Y(q,\tau)$ は次式のように分解して表すこともできる。

$$\begin{aligned} &C_{Y,A,\mathrm{min}}(q,\tau) + C_{Y,\phi,\mathrm{min}}(q,\tau) + C_{Y,\phi,\mathrm{all}}(q,\tau) \\ &= C_{\mathrm{src,A,min}}(q,\tau) + C_{\mathrm{src},\phi,\mathrm{min}}(q,\tau) + C_{\mathrm{src},\phi,\mathrm{all}}(q,\tau) \\ &\quad + C_{\mathrm{flt,A,min}}(q,\tau) + C_{\mathrm{flt},\phi,\mathrm{min}}(q,\tau) + C_{\mathrm{flt},\phi,\mathrm{all}}(q,\tau) \\ &\quad + C_{H,A,\mathrm{min}}(q,\tau) + C_{H,\phi,\mathrm{min}}(q,\tau) + C_{H,\phi,\mathrm{all}}(q,\tau) \end{aligned} \qquad (5)$$

式 (4) に従えば，残響環境下での F_0 推定の課題は，フィルタ特性と残響特性を取り除くことで，$C_Y(q,\tau)$ から $C_{\mathrm{src}}(q,\tau)$ のみを推定することである。しかし，$h(t)$ あるいは $C_H(q,\tau)$ を測定することなく，$C_{\mathrm{src}}(q,\tau)$ のみを取り扱うことは非常に難しい。また，分析幅が残響時間よりも長い状態の $C_H(q,\tau)$ が，正確な $C_{\mathrm{src}}(q,\tau)$ を抽出するためには必要である。

3.3 提案法

図 1 に，複素ケプストラム分析を利用し，MTF の概念と音源フィルタモデルの概念を適用して実現した F_0 推定法のアルゴリズムを示す。この方法は，大まかに，次のような三つのフェーズで表される。

(A) 残響インパルス応答の推定と残響音声からの主要残響成分（推定の妨害になるもの）の除去

(B) 音源フィルタモデルに基づいたリフタリングを利用してフェーズ (A) の処理を施された残響音声から，音源情報 $X_{\mathrm{src}}(\omega,\tau)$ を推定する処理

(C) 上記の特徴から最終的に F_0 を推定する処理

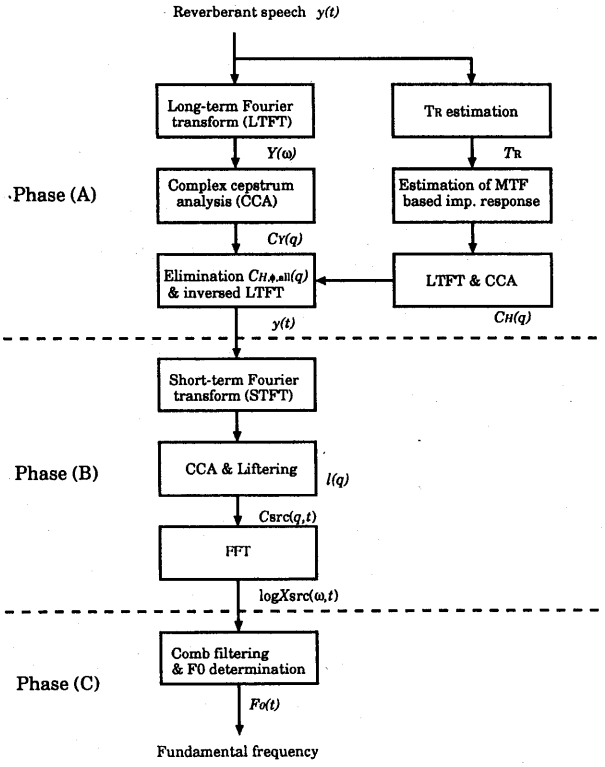


図 1　提案法のアルゴリズム
Fig. 1 Algorithm for proposed method

図 1 に示すように，最終決定ブロックとして，本稿では，Comb フィルタリング [18] を採用した。また，図 2 に示すリフタリングは

$$l(q) = \begin{cases} 1, & q > q_{\mathrm{lif}} \\ 0, & q \leq q_{\mathrm{lif}} \end{cases} \tag{6}$$

とした。但し，$q_{\mathrm{lif}} = 1.25$ ms である。F_0 推定値の下限と上限は，それぞれ 80 Hz と 800 Hz とした（詳細な記述については文献 [12, 13] を参照のこと）。

この提案法は，次に示すような著者の研究グループの検討結果 [12, 13] に基づいて得られたものである。

(1) 残響インパルス応答 $h(t)$ の all-pass 位相成分は，最小位相成分と比較すると，ロバストで正確な F_0 推定に悪影響を与える主要な原因である。そのため，LTFT で表現された式 (5) において，$C_{H,\phi,\mathrm{all}}(q,\tau)$ を取り除くことで $h(t)$ による影響を低減させることができる。

(2) MTF の概念に基づくことで，残響がどれだけ MTF の減衰に影響を与えているか知ることができる。そのため，逆 MTF の特性を利用することで室内伝達特性（残響時間 T_R）を予測することができる。

(3) $h(t)$ が指数関数的に減衰する形で定義されたと仮定すると，$C_{H,A}(q,\tau)$ もケフレンシーに関して指数関数的に減衰することが容易に推測できる。そこで，統計的な近似として模擬された白色雑音キャリア $\hat{n}(t)$ と $\hat{a}\exp(-6.9t/\hat{T}_R)$ を利用することで見掛け上，$h(t)$ を推測することができる。

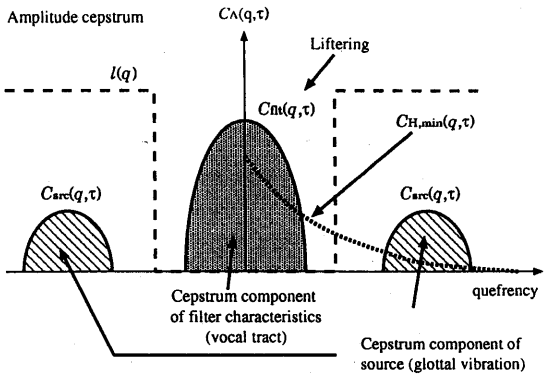


図 2　ケフレンシー領域における音源特性とフィルタ特性の表現
Fig. 2 Separated representations of source and filter characteristics in quefrency domain

(4) $C_{A,\min}(q,\tau)$ と $C_{\phi,\min}(q,\tau)$ の間には，Hilbert 変換の関係があり，$C_{H,\phi,\min}(q,\tau)$ は，最小位相特性に基づくため正のケフレンシーで同じ特性をもつ。低ケフレンシーにおける $C_{H,\min}(q,\tau)$ は，高ケフレンシーのものよりも一般に大きく，この値はケフレンシーの増加とともに指数関数的に減衰する（図 2 に点線で示す）。そのため，$C_{H,\min}(q,\tau)$ は，低ケフレンシー部に集中するものと仮定できる。

(5) 音源フィルタモデルに基づくと，音声の音源特性とフィルタ特性のケプストラム成分を，それぞれ，高ケフレンシー部と低ケフレンシー部に分けて表すことができる。そのため，もし低ケフレンシー部の成分だけがリフタリングにより取り除かれたとすれば，$C_{\mathrm{flt}}(q,\tau)$ と $C_{H,\min}(q,\tau)$ は，式 (5) で除去されることになる（図 2 のリフタリング処理を参照）。

3.4　F_0 推定方法の例

一例として，図 3 に，式 (5) を利用して $C_{\mathrm{src}}(q,\tau)$ から F_0 を推定する際，$C_{H,\min}(q,\tau)$ と $C_{H,\mathrm{all}}(q,\tau)$ のどちらの成分が F_0 推定に悪影響を与えるのか調査した結果を示す。ここで用いた音声信号は，女性話者の/Tokushima-To-Ieba-Awa-Odori-Ga-Yuumei-Desu/であり，残響時間 $T_R = 2.0$ s の人工残響を畳み込んで得られた残響音声 $y(t)$ を対象とした。図 3(a)，図 3(b) に，それぞれ原音声 $x(t)$ と残響音声 $y(t)$ を示す。各パネルの破線すべては，F_0 の正解値（詳細は，3.2 節で説明する）を示す。図 3(c) の実線で示される $F_{0,\mathrm{Est}}(t)$ は，$y(t)$ からケプストラム法を利用して推定されたものである。この結果から，F_0 推定値は参照 F_0 に対し，全体的にずれ，一致していないことがわかる。しかし，図 3(d) に示すように，LTFT の複素ケプストラム上で $h(t)$ の成分を $y(t)$ から取り除いた後，この方法を適用すれば正確に F_0 を推定できることがわかる。ここで，

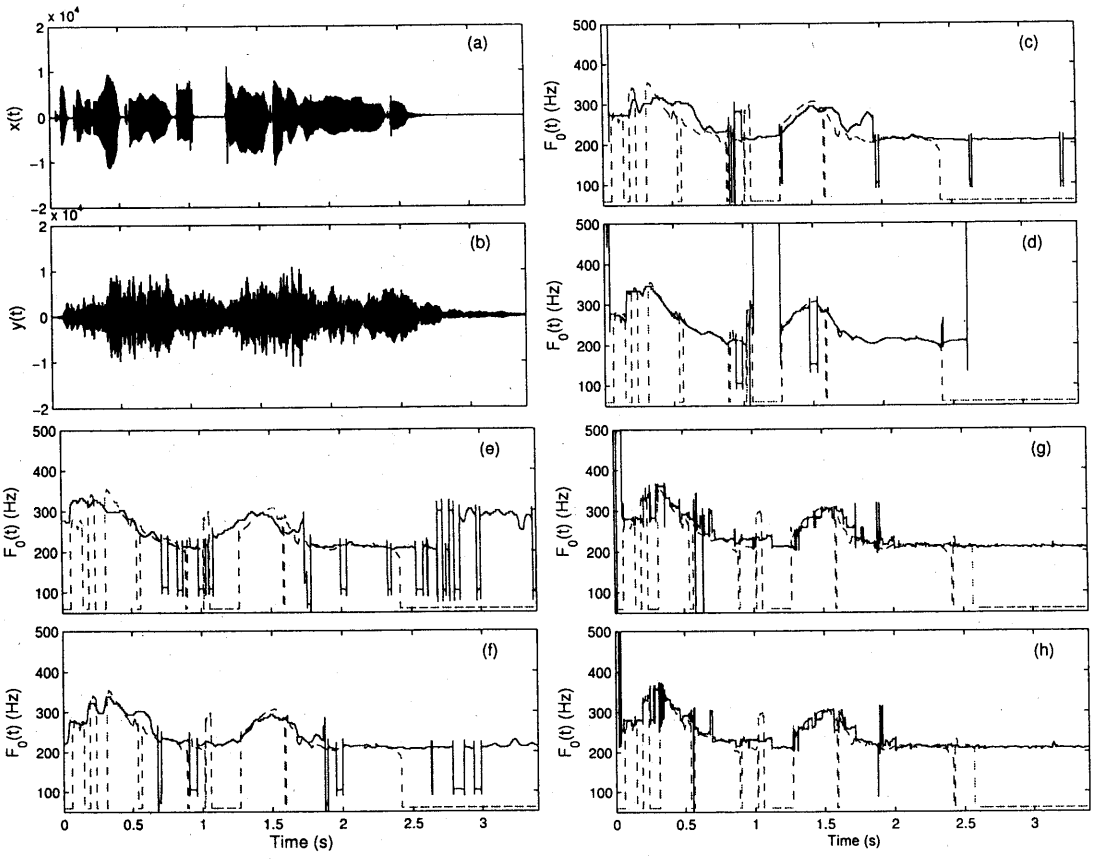


図 3 F_0 推定例: (a) 原音声 $x(t)$, (b) 残響音声 $y(t)$ ($T_R = 2.0$ s), (c) 参照 F_0 (破線) とケプストラム法による F_0 推定 (実線), (d) $h^{-1}(t)$ を利用して残響除去された $y(t)$ からの F_0 推定, (e) 最小位相成分を除去した $y(t)$ からの F_0 推定, (f) 非最小位相成分を除去した $y(t)$ からの F_0 推定, (g) T_R 推定なしと (h) T_R 推定ありの提案法による F_0 推定

Fig. 3 Example of the F_0 estimation: (a) original signal $x(t)$, (b) reverberant speech $y(t)$, ($T_R = 2.0$ s), (c) reference F_0 (dashed-line) and estimated F_0 from $y(t)$ using the cepstrum method (real-line), (d) estimated F_0 from $y(t)$ using $h^{-1}(t)$, (e) F_0 from $y(t)$ eliminated by minimum phase characteristics, (f) F_0 from $y(t)$ eliminated by all-pass phase characteristics, (g) estimated F_0 using proposed method without T_R estimation, and (h) with T_R estimation

$h(t)$ の最小位相成分ならびに非最小位相成分のみを取り除いたものに対し，推定法を適用すると，図 3(e), 図 3(f) に示すように，一方では推定が失敗し，一方では推定が成功していることがわかる。

4. 基本周波数推定の比較評価

4.1 F_0 推定の従来法

本稿では，9 個の代表的な F_0 推定法がどれだけ残響環境下でロバストに推定できるかを評価する。ここでは，AMDF 法 [4]，短時間 Fourier 変換における自己相関法 (STFT-ACorrLog: AutoCorrelation of Log-amplitude spectrum) [4]，短時間 Fourier 変換における Comb フィルタリング法 (STFT-Comb: Comb filtering)[4]，SHS (sub-harmonic summation) 法 [4]，ケプストラム (Cepstrum) 法 [6]，LPC 残差 (LPC-residue) 法 [4]，TEMPO [8]，YIN [9]，PHIA (Periodicity/Harmonicity using Instantaneous Amplitude) 法 [11] を利用した。他に，ここに述べたもの以外の手法も評価に利用したが，それらの多くは，ここに取り上げた手法の改良法や類似した方法であるため，本稿では上記にあげた 9 個を比較対象に限定した。提案法に関しては，T_R 推定を利用するもの ("Prop(Est)") としないもの ("Prop(Org)") の 2 種類を評価に利用した。T_R 推定を利用しない場合は，T_R を既知として推定精度を評価した。これは，T_R 推定がどれだけ正確にできているか，また T_R 推定の誤差がどれだけ F_0 推定の精度に影響を与えるか調べるためのものである。また，もう一つの比較法として，音源フィルタモデルに基づく複素ケプストラム上の F_0 推定法を利用した ("SrcFlt") [18]。この方法は，$C_{H,\phi,\mathrm{all}}(q,\tau)$ が LTFT 上で取り除かれることで，どれだけ推定精度の向上に効果があるかをみるためのものである。

4.2 音声データセットと評価尺度

この評価で利用したデータは，Atake らによって収録された音声データセットである [10]。このデータセットは，男性 14 名，女性 14 名の日本語文章（30 文章）を発話音声と EGG を同時収録したもので構成されたものである（合計 840 発話，16 kHz サンプリング，16 bits 量子化）。

評価実験では，原音声 $x(t)$ に次式で定義される残響インパルス応答（Schroeder の人工残響インパルス応答 [1, 2]）$h(t)$ を畳み込むことで得られた残響音声 $y(t)$ を利用する。

$$h(t) = a \exp\left(\frac{-6.9t}{T_R}\right) n(t) \tag{7}$$

但し，a は $h(t)$ の正規化ゲイン項，T_R は，T_{60} に基づく残響時間，$n(t)$ は白色雑音である。この人工的な残響インパルス応答は，非最小位相特性を有するものであり，統計的室内音響学でよく利用されるものである（既に第 1，第 2 稿 [1, 2] で紹介したものである）。ここでは，6 種類の残響時間（T_R = 0.0, 0.1, 0.3, 0.5, 1.0, 2.0 s）を利用する。そのため，実験で利用する残響音声の総数は 5,040 個（840 発話 ×6 条件）である。

次に，実残響環境における音声信号についても，同様に実残響インパルス応答を利用して，原音声に畳み込むことで残響音声を得た。ここでは，30 種類の実残響インパルス応答（残響データベース SMILE [20]）を利用した。そのため，実験で利用する残響音声の総数は，25,200 個（840 発話 ×30 条件）である。

本稿では，F_0 推定のロバスト性と正確性を検討するために，次式で定義される正答率（Correct rate）（%）を評価尺度として利用した。

$$\text{Correct rate} = \frac{N_{F_{0,\text{Est}}}(E)}{N_{F_{0,\text{Ref}}}} \times 100 \tag{8}$$

但し，$N_{F_{0,\text{Est}}}(E)$ は，$F_{0,\text{Ref}}(t)$ と $F_{0,\text{Est}}(t)$ をそれぞれ参照 F_0（正解）と推定 F_0 としたときの，有声区間内で $|F_{0,\text{Ref}}(t) - F_{0,\text{Est}}(t)|/F_{0,\text{Ref}}(t) \leq E(\%)$ を満たす正答数である。$N_{F_{0,\text{Ref}}}$ は有声区間内の $F_{0,\text{Ref}}(t)$ の数を示す。本稿では，目的となる原音声の EGG 信号から，TEMPO で推定された F_0 を $F_{0,\text{Ref}}(t)$ とした。ここで正答率については，先の検討報告に習い，推定誤差（$E = 5$ %）以内の正答率を求めた。

4.3 評価結果

図 4 に，残響時間 T_R を関数とした，代表的な推定法ならびに提案法による残響音声からの F_0 推定の比較結果を示す。この図は，F_0 推定における推定誤差（$E = 5$ %）内の正答率（%）を示している。代表的な方法の正答率は，残響時間 T_R の増加とともに，劇的に減少していることがわかる。特に，残響時間が 2.0 s のとき，正答率は 50 %未満となった。全体的に，従来法では，残響時間が増加するにつれ F_0 の推定精度が低下する傾向にあるが，提案法では，これらの結果に対して約 10 %の改善がみられることがわかる。

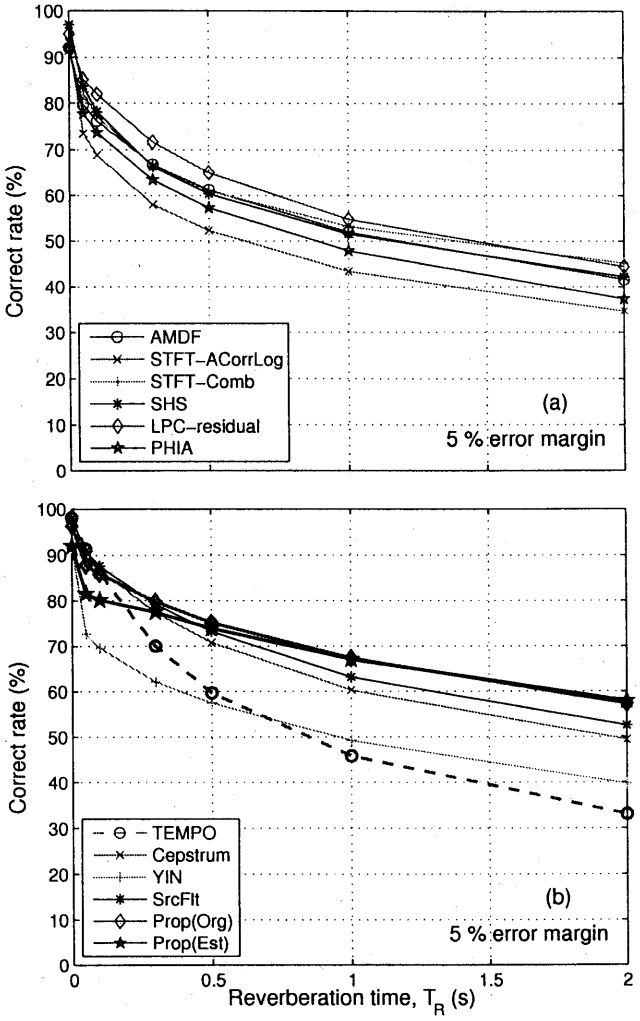


図 4 人工的な残響環境における推定誤差 5 %内の正答率。残響時間 T_R の関数とした残響音声の F_0 推定結果

Fig. 4 Estimation results: percent correct rate within error margin of 5 % of F_0 estimates from reverberant speech as function of T_R

提案法の結果では，T_R 推定を行う場合と行わない場合でそれほど大きな違いはみられていないことから，T_R 推定がよく機能しているものと考えられる。提案法では，残響時間 T_R = 2.0 s で 60 %の正答率を得ていることから，T_R 推定を利用することで，MTF ベースの残響インパルス応答を正確に推定できているといえる。もう一つの SrcFlt 法の結果は，ケプストラム法の結果より若干の改善（正答率で約 3 %の改善）を示している。対照的に，提案法を利用することでケプストラムの結果よりも約 7 %の改善を得ている。このことから，複素ケプストラムの利用（最小位相と非最小位相成分の切り分け）が，残響環境下で F_0 推定のために効果的であるといえる。

例えば，図 3(b) の残響音声 $y(t)$ に対し，T_R 推定あり・なしの提案法の推定結果を図 3(g), (h) に示す。この結果からも，提案法が残響に関してロバストでかつ

正確に F_0 を推定できていることがわかる。

提案法といくつかの代表的な方法（TEMPO，YIN，ケプストラム法，SrcFlt）について，実残響インパルス応答を利用した比較評価実験の結果（すべての条件下で平均した正答率）を表 1 に示す。他の手法に関する結果は，劇的な改善を示していないため，ここには示していない。提案法 Prop(Org) によって得られた結果のほとんどが，全体の中でベストな結果となった。この表は，残響時間 T_R が正確に推定されるときに，提案法 Prop(Est) が非常によく機能することを示している。提案法によって得られた改善は，TEMPO の結果に対して 20 %以上，ケプストラム法の結果に対して 10 %以上になった。これは，提案法の推定アルゴリズムが，残響に関するロバスト性に関する問題の解決策として利用できるものと考えられる。

5. 残響時間推定法の取組み

これまで概説してきたように，残響時間は系の残響特性を表す重要なパラメータの一つであるため，特定の残響特性を有する室の設計や評価，目的に応じた室の選別に利用される [14]。また，残響音声回復法や残響環境下の音声認識システム，本稿前半で概説したような基本周波数推定法でも，重要なパラメータの一つとして利用されている [2, 12, 13].

残響時間は，室内に放射された駆動音源が平衡状態に達した後，信号が止められた時刻からそのエネルギー密度の平均減衰曲線が 60 dB 低下する時刻（T_{60}）として定義される [14, 15]。この減衰曲線は，インパルス積分法 [15] を利用して，インパルス応答から求められるため，正確な残響時間を知るためには，安定で高精度な系の伝達特性（室内残響インパルス応答）の測定が求められる。そのため，ピストルの発砲音といったインパルス性の音源信号を利用した直接観測法 [14] や TSP 信号を利用した算出法 [16] が考えられてきた。後者の方法は，音源の特性と測定の安全面の点から，現在でも精密な残響時間測定に利用されている。

しかし，実際に我々が利用するような残響環境下で残響時間を測定しようとする場合，こういった精密な測定ができる状況がいつでも与えられるとは限らない。例えば，測定装置の適切な配置の問題や，測定の妨げとなる喧噪への対応が考えられる。また，一般的に静寂な環境と思われる室内ですら 30〜40 dB SPL 相等の暗騒音が存在する。実際，T_{60} の算出には 60 dB 以上のダイナミックレンジを必要とすることから，こういった環境では 100 dB SPL 以上の非常に大きなインパルス性の音源信号を利用して，系の伝達特性を測定しなければならない。実際の室の特性の精密測定を考えると，室内に人がいたりする場合があるため，こういった測定はほぼ不可能といえる。

一方，前述したように，残響時間は様々な応用音声信号処理（例えば，残響音声回復）で利用されはじめている。これらの応用信号処理が，定常な特性をもつ室ではなく，むしろ人や物が動き，温度や湿度が時々刻々と変化するような環境で利用されることを考えると，これらの変化とともに時々刻々と変化する残響時間を随時測定して，応用信号処理に受け渡さなければいけない。測定機器の準備や設置の必要性，残響時間の算出にかかる所要時間を考慮しても，測定系と応用信号処理を同時に行う必要性があることから，このような逐次的な装置の組合せは現実的であるとはいえない。

以上のことから，系の伝達特性を測定せずに，残響時間を知る方法を実現することができれば，系の特性が変動したとしても残響時間を知ることができるかもしれない。例えば，室内で観測される音のみを利用して残響時間を知ることができれば，測定環境上の制約（測定装置の設置，室内の人や物による配置の影響，室の状況（温度・湿度）による影響）をほとんど無視することができる。

本稿では，著者の研究グループで既に検討した残響時間のブラインド推定法を紹介する。ここでは，室内で観測された音声（音）信号のみを利用して，残響時間をブラインド推定できるかどうか，その可能性について検討するとともに，条件付ではあるが，精度よく残響時間をブラインド推定できる方法を概説する。

6. MTF に基づいた残響時間推定法

6.1 従来法とその問題点

第 1 稿 [1] で，MTF に基づいたパワーエンベロープ逆フィルタ処理を概説した（第 2，第 3 節，式 (1)〜(24)）。このときの重要な関係式を下記にまとめる。

$$e_y^2(t) = e_x^2(t) * e_h^2(t), \quad t = nT_s \tag{9}$$

$$\begin{aligned} \mathbf{Z}[e_x^2[n]] &= \mathbf{Z}[e_y^2[n]]/\mathbf{Z}[e_h^2[n]] \\ &= \frac{\mathbf{Z}[e_y^2[n]]}{a^2} \frac{1}{1-\exp\left(-\frac{13.8}{T_R \cdot f_s}\right) z^{-1}} \end{aligned} \tag{10}$$

但し，f_s はサンプリング周波数（20 kHz），n は離散時間サンプル（$t = nT_s$，$T_s = 1/f_s$）である。

式 (2) に示すように，残響信号 $y(t)$ は，原信号 $x(t)$ と室内残響インパルス応答 $h(t)$ の畳み込みとして表現されるが，それぞれのパワーエンベロープ $e_y^2(t)$，$e_x^2(t)$，$e_h^2(t)$ に関して整理すると，式 (9) のようにパワーエンベロープ上での畳み込みが成り立つことがわかっている [1, 2]。ここで，$h(t)$ が式 (7) のように Schroeder の近似式であるとすると，周波数領域では簡単な乗算の関係から，$e_x^2(t)$ を求めることができる。本稿では，離散信号として算出するため，各パワーエンベロープは $e_y^2[n]$，$e_x^2[n]$，$e_h^2[n]$ と表され，それらの周波数領域での関係は，式 (10) のように z 変換（$\mathbf{Z}[\cdot]$）を利用して表現される。以上のことから，$e_x^2(t)$ は，最終的に式 (10) を逆 z 変換することで得られる。

MTF ベースのパワーエンベロープ逆フィルタは，二つのパラメータ（a と T_R）によって決まる。ここで，パラメータ a はゲイン項であるため，パワーの正規化として定められるが（第 1 稿 [1]，式 (24) 参照），パラメータ T_R は次式によって推定可能であることが既に明らかにされている。

$$\hat{T}_R = \max\left(\arg\min_{T_R} \int_0^T |\min(\hat{e}_{x,T_R}^2(t), 0)| dt\right) \tag{11}$$

表 1　実際の残響環境における推定誤差 5 %内の正答率の比較。IRdata は，データファイル [20] の番号に対応する。残響時間 T_R は 125 Hz から 8 kHz のオクターブ周波数での伝達関数に対する残響時間の平均値を示す。ボールド体とイタリック体は，一番良い結果と悪い結果を示す。

Table 1 Comparison of percent correct rate (%) within error margin of 5 % in actual reverberant environments. IRdata corresponds to File No. in [20]. Reverberation time, T_R, is the average of T_Rs for transfer functions at 125 Hz to 8 kHz at octave frequencies. Bold and italic faces indicate best and worst results

室の条件 (RIR)	IRdata	残響時間 T_R (s)	TEMPO	YIN	Cepstrum	SrcFlt	Prop(Org)	Prop(Est)
クリーン	—	0.00	**98.10**	91.49	96.65	95.81	95.81	*90.69*
多目的ホール 1 [1]	301	1.09	*28.89*	35.51	42.22	40.62	**50.70**	44.99
多目的ホール 1 [2]	302	0.80	*33.13*	38.36	47.54	47.04	**56.79**	51.67
多目的ホール 2 [3]	303	1.44	*22.77*	30.27	38.35	35.10	**43.58**	38.96
多目的ホール 2 [4]	304	1.04	*32.08*	37.22	47.10	46.47	**55.02**	51.12
多目的ホール 3 [5]	305	1.93	*19.84*	27.91	32.70	31.07	**41.62**	36.52
多目的ホール 3 [6]	306	1.35	*26.62*	33.47	39.51	39.41	**50.89**	44.70
多目的ホール 4 [7]	307	1.42	*30.21*	38.21	46.89	49.52	**60.54**	54.50
多目的ホール 4 [8]	308	1.54	*29.44*	37.21	46.04	49.01	**60.04**	54.20
多目的ホール 5 [9]	319	1.47	*32.37*	39.25	48.55	48.98	**56.31**	53.29
多目的ホール 6 [10]	321	2.16	*27.64*	35.05	44.69	46.75	**53.41**	50.37
コンサートホール 1 [11]	309	2.35	*23.42*	29.37	35.49	37.64	**44.30**	43.06
コンサートホール 1 [12]	310	2.34	*29.61*	33.71	43.13	46.93	49.59	**50.37**
コンサートホール 1 [13]	311	2.35	*24.90*	29.85	37.51	41.02	**48.92**	45.66
コンサートホール 1 [14]	312	2.39	*18.26*	26.07	32.70	31.55	**40.72**	35.46
コンサートホール 1 [15]	313	2.38	*14.73*	22.97	29.68	27.32	**37.38**	29.76
コンサートホール 2 [16]	314	1.14	*24.19*	32.23	38.12	36.11	**45.18**	40.51
コンサートホール 3 [17]	315	1.96	*25.94*	35.72	**46.85**	43.04	44.27	45.33
コンサートホール 4 [18]	316	1.92	*23.53*	31.26	38.46	38.64	**50.82**	42.81
コンサートホール 4 [19]	317	2.55	*20.47*	27.23	34.49	36.13	**48.09**	40.33
コンサートホール 5 [20]	323	2.32	*25.41*	33.39	41.95	41.79	**48.20**	45.70
コンサートホール 6 [21]	324	1.77	*29.99*	36.07	45.29	47.20	**53.82**	51.61
コンサートホール 6 [22]	325	1.74	*34.16*	38.28	47.13	49.79	**55.87**	54.07
コンサートホール 6 [23]	326	1.69	*18.38*	23.61	27.84	28.55	**41.62**	31.53
教室 [24]	201	1.36	*32.56*	41.51	53.50	51.48	**57.89**	55.36
劇場 [25]	318	0.85	*32.90*	38.06	46.39	45.05	**54.16**	50.09
会議室 [26]	401	0.62	*57.04*	55.28	70.26	70.25	**72.58**	71.14
教室 [27]	402	1.12	*36.74*	47.03	61.52	56.74	**61.78**	60.14
教室 [28]	403	1.09	*26.48*	35.57	44.59	42.22	**52.71**	46.30
スピーチホール 1 [29]	404	1.54	*23.34*	31.97	40.04	38.11	**47.47**	41.71
教会 1 [30]	405	0.71	*32.46*	38.97	47.31	43.66	**52.27**	48.22
教会 2 [31]	406	1.30	*23.67*	30.32	36.84	36.15	**45.29**	41.91
イベントホール 1 [32]	407	3.03	*16.99*	22.81	26.91	27.23	**37.68**	31.94
イベントホール 2 [33]	408	3.62	*15.19*	21.78	26.38	27.14	**37.61**	29.68
体育館 1 [34]	409	2.82	*19.19*	25.95	31.25	32.81	**44.95**	35.07
体育館 2 [35]	410	1.70	*22.35*	27.77	31.70	32.67	**45.67**	36.08
リビングルーム [36]	411	0.36	74.24	*65.35*	**81.45**	73.40	72.08	69.72
映画館 [37]	412	0.38	*42.88*	43.30	52.16	51.96	**59.32**	56.85
アントリウム [38]	413	1.57	*27.58*	30.65	34.56	38.00	**50.69**	44.01
トンネル [39]	414	2.72	*14.03*	22.43	25.30	24.88	**33.06**	28.54
駅のコンコース [40]	415	1.95	*20.89*	24.79	27.57	29.88	**44.71**	36.44
スピーチホール 2 [41]	416	1.53	*29.15*	36.29	45.74	45.78	**54.54**	49.79
スピーチホール 2 [42]	417	1.49	*22.17*	30.49	37.58	37.04	**47.07**	41.13
スピーチホール 2 [43]	418	1.40	*18.96*	25.71	29.46	27.83	**40.22**	30.72

[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

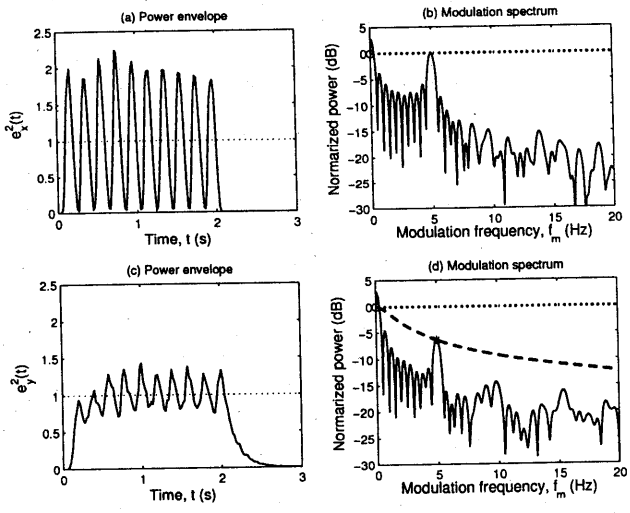


図 5　抽出されたパワーエンベロープ（(a) と (c)）と変調スペクトル ((b) と (d))。上段は $T_R = 0.0$ s, 下段は $T_R = 2.0$ s の場合

Fig. 5 Extracted power envelopes ((a) and (c)) and modulation spectra ((b) and (d)) of reverberant sinusoids in cases of $T_R = 0.0$ s (upper panel) and $T_R = 2.0$ s (bottom panel)

ここで T は信号時間長，$\hat{e}^2_{x,T_R}(t)$ は残響時間 T_R の関数として回復されたパワーエンベロープである。この式は，回復された $\hat{e}^2_x(t)$ の最も深い谷が 0 に達するときに回復が完全になされたとして，T_R を推定するものである。

先に述べたように，$\hat{T}_R$ は最善のパワーエンベロープ回復の意味で最適に求められるものであるが，この推定値が系の残響時間 T_R の値よりも過小推定され，この不一致が元々の残響時間 T_R が増加するにつれ大きくなる傾向にあることがわかっている（詳しくは次節で説明するが，図 8 と図 9 の破線を参照）。これは，残響時間のブラインド推定法として，従来法（式 (11)）を利用するには問題があることを意味している。

この問題が起こる原因は，残響信号から抽出されたパワーエンベロープを回復する際，LPF で取りきれなかった高周波成分がパワーエンベロープ逆フィルタ処理（微分処理）で強調され，それらの位相の状態によっては，残響時間推定の精度を左右する谷の形成（変調度を定義するもの）に影響を与えるためである。この谷の形成は，もともとパワーエンベロープを構成する主要な変調周波数成分が作るものよりも点在することになるため，式 (11) より，残響時間が過小に推定される結果となっている。

6.2　推定法の概念

図 5 に，変調周波数 5 Hz の人工的な信号に対するパワーエンベロープ（図 5(a) と図 5(c)）とそのときの変調スペクトル（図 5(b) と図 5(d)）を示す。図の上段は残響がない場合，下段は $T_R = 2.0$ s の残響が付与された場合である。図 5(b) と (d) に描かれた信号の変調

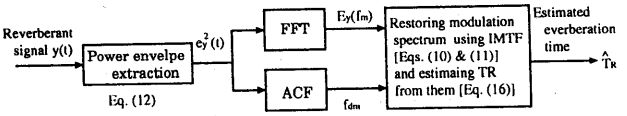


図 6　残響時間推定のブロックダイアグラム

Fig. 6 Block diagram for estimating reverberation time

スペクトルを見ると，次の三つの特性があることがわかる。

(1) 変調周波数 0 Hz（DC）の成分が残響の影響を受けていないこと（0 dB，図中の “o” 印）

(2) 残響のない状態の変調スペクトルの主要な変調周波数のパワーは 0 Hz でのものと同じ値にあること（0 dB，変調度 1）

(3) 変調周波数 5 Hz での変調スペクトルは残響が付加されると MTF に従い減少すること

ここで見られる特徴は，様々な変調周波数（$f_m \leq 20$ Hz）で観測されている。

これらの有用な特性は，観測した信号から残響時間をブラインド推定するための方法のモデル化を可能にさせてくれる。これは，明確な残響時間が主要な変調周波数 f_{dm}（例えば，図 5 における $f_{dm} = 5$ Hz）での減少した変調スペクトルを補償すること（$m(f_m)$ が 1 に回復されること）で残響時間 T_R を推定可能であることを意味している。従来法（式 (11)）では，時間領域で減少した変調度からそれが 1 に戻るように信号のパワーエンベロープを回復することで処理されたのに対し，ここで述べた方法は，変調周波数領域でパワーエンベロープの変調度に関係した変調スペクトルを取り扱うことで，安定で正確な残響時間推定を可能とする。

6.3　提案法

6.1 節に示した推定法のコンセプトに基づき，変調周波数領域で残響時間をブラインド推定する方法を提案する。そこで，提案法では先の二つの特性に基づき，

$$\log|E_x(f_{dm})| = \log|E_x(0)| \quad (12)$$

$$\log|E_y(0)| = \log|E_x(0)| \quad (13)$$

と仮定する。但し，$E_x(f_m)$ と $E_y(f_m)$ は，それぞれ，$e^2_x(t)$ と $e^2_y(t)$ の変調スペクトルである。式 (12) と式 (13) は，特性 (1) と (2) を表したものである。これらは提案法での最初の仮定であるが，実際上，これらが有用な特性であることを確認している。これらに基づくと，残響時間の推定値 $\hat{T}_R$ は，特性 (3) によって，次式のように減少した変調スペクトルから推定できる。

$$\hat{T}_R = \arg\min_{T_R} \left(\left|\log|E_y(f_{dm})| - \log|E_y(0)| - \log\hat{m}(f_{dm}, T_R)\right|\right) \quad (14)$$

但し，$\log|E_y(f_{dm})| - \log|E_y(0)|$ は特定の変調周波数 f_{dm} での減少した変調スペクトル，$\hat{m}(f_{dm}, T_R)$ は，T_R

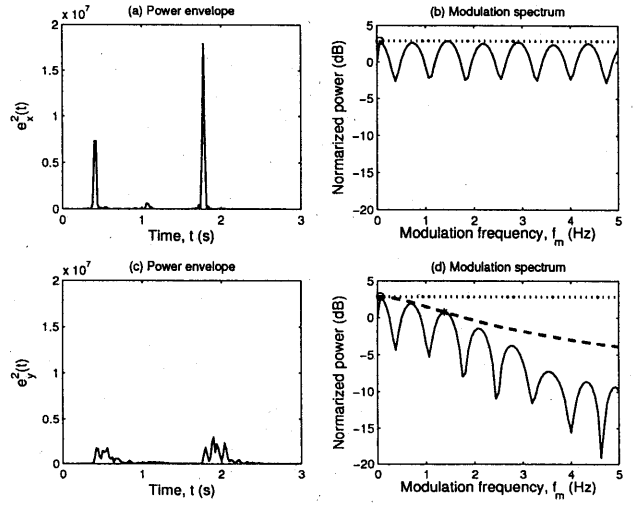


図 7　抽出されたパワーエンベロープ（(a) と (c)）と変調スペクトル (b) と (d))。上段は $T_R = 0.0$ s, 下段は $T_R = 2.0$ s の場合

Fig. 7 Extracted power envelopes ((a) and (c)) and modulation spectra ((b) and (d)) of reverberant speech in cases of $T_R = 0.0$ s (upper panel) and $T_R = 2.0$ s (bottom panel)

の関数として f_{dm} での MTF である。この式は，$m(f_{dm})$ が 1 に回復されるときの T_R を決めることに相等する。

図 6 は，式 (14) を利用して残響時間 T_R をブラインド推定する方法のブロックダイアグラムを示す。ここで，FFT は高速 Fourier 変換，ACF は自己相関関数である。ACF は，パワーエンベロープ $e_y^2(t)$ に対し，変調スペクトル $E_y(f_m)$ での主要な変調周波数 f_{dm} を求めるために利用される。パワーエンベロープを回復する意味で，式 (10) のパワーエンベロープ逆フィルタが $\hat{m}(f_{dm}, T_R)$ を得るために利用された。

例えば，図 5(d) の破線は，提案法で導出された残響時間 $\hat{T}_R$ を示す。図 7(a) と図 7(c) は帯域制限された信号それぞれのパワーエンベロープを示し，図 7(b) と図 7(d) は，それらに対応する変調スペクトルを示す。図 7 のフォーマットは，図 5 のものと同じである。図 5(a) のパワーエンベロープにおいて，主要な変調周波数（f_{dm} Hz）での変調スペクトルが 0 Hz 近くでのものと同じ値になっている。図 7 で示されたようなパワーエンベロープは，よく帯域制限された音声信号で見られるものである。

7. 残響時間推定の比較評価

7.1 モデルの検証評価：人工的な残響環境の場合

本稿では，提案法が原理どおりに残響時間をブラインド推定できるかどうかを検証するために，残響音声信号を利用して提案法を評価する。ここでは，図 5(a) に示すようなパワーエンベロープをもつ人工的な AM 信号 $x(t)$（変調周波数を 5 Hz とする正弦波を雑音キャリアに振幅変調したもの）と，女性話者（faf）によって

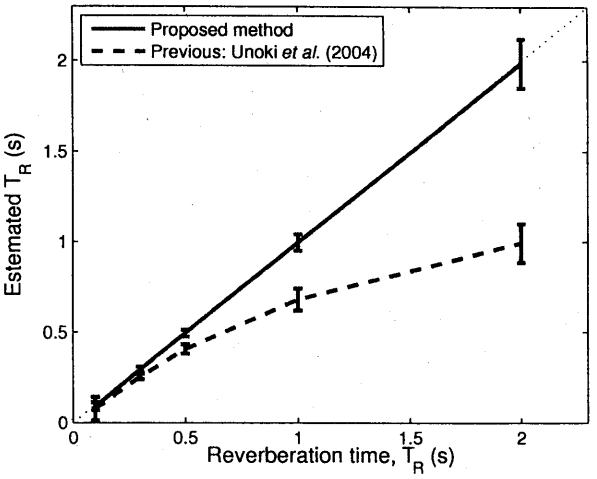


図 8　残響が付加された人工信号音からの残響時間の推定結果

Fig. 8 Estimated reverberation time from reverberant sinusoids

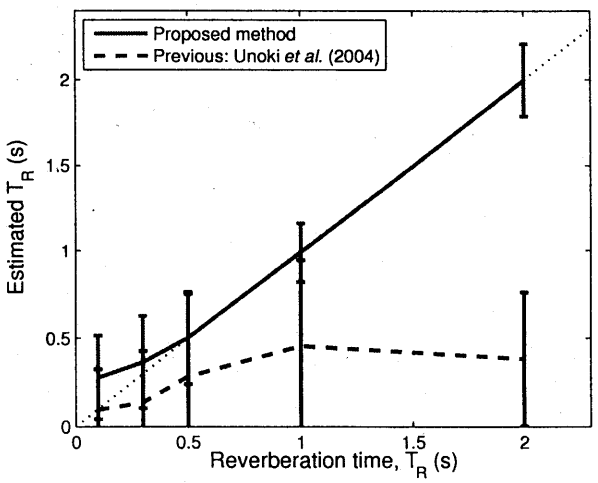


図 9　残響音声からの残響時間の推定結果

Fig. 9 Estimated reverberation time from reverberant speech

発話された日本語文章の音声信号 $x(t)$ [19] に対し評価した。一つの残響時間当たり 100 個の人工的な室内インパルス応答（式 (7) の $h(t)$）として，合計 5 種類の残響時間（$T_R = 0.1, 0.3, 0.5, 1.0, 2.0$ s）に対して評価した。すべての音声信号は，定帯域フィルタバンク（100 Hz 帯域幅で 100 チャンネル分割）を利用して分解される。推定対象となるチャンネルの選択は，提案法を音声信号に適用させられるようにするために，いくつかの制約を持つことになるが，ここでは，音声のスペクトル成分が密にあるところを推定対象のチャネルとするように設定した。すべての残響信号 $y(t)$ は，人工的な信号の場合で 500（$= 100 \times 5$）個，音声信号の場合で 4,000（$= 100 \times 5 \times 8$）個となる。

図 8 と図 9 は，人工的な残響 AM 信号（図 5）から推定された残響時間 $\hat{T}_R$ と，音声信号（図 7）から推定された残響時間 $\hat{T}_R$ を示す。図中の点は $\hat{T}_R$ の平均値を，

エラーバーは標準偏差を示す。図中の点線は，理想推定値を示し，破線は従来法で推定された値を示す。両図で，従来法によって推定された残響時間 $\hat{T}_R$ は，T_R が増加するにつれ，過小推定される傾向にある。これに対し，図8の場合はすべてに渡って，推定値 $\hat{T}_R$ が理想値 T_R に一致し，図9の場合は $T_R = 0.3$〜2.0 s の範囲で一致している。図9では，残響音声に対して数チャンネルでの T_R 推定を利用した場合，推定値 $\hat{T}_R$ に対する標準偏差が減少する傾向にあった。

7.2 応用評価：現実の残響環境の場合

次に，提案法が，実際の残響環境で基本的なコンセプトに基づいて残響時間 T_R をブラインド推定できるかどうか調べるために，残響音声信号を利用して提案法を評価する。ここでは，SMILE データベース [20] にある17個の室内残響インパルス応答を利用した。これらの室内インパルス応答は，木造住宅のリビングルーム（$T_R = 0.36$ s），映画館（0.38 s），会議室（0.62 s），教会（0.71 s），9個の多目的ホール（0.80, 1.04, 1.09, 1.35, 1.42, 1.47, 1.54, 1.93, 2.16 s），シアターホール（0.85 s），3つのクラシックコンサートホール（1.69, 1.77, 1.96 s）であった。また，上記の評価と同様の音声データ [19] を利用した。図10は，17個の残響音声信号から残響時間 T_R を推定した結果を示す。相対的に残響時間が短い（≤ 1.0 s）場合，ブラインド推定された残響時間 $\hat{T}_R$ は，かなり安定で正確な値を示した。しかし，残響時間が長い場合，ブラインド推定された $\hat{T}_R$ は過小推定値を示した。三つのラベル（A, B, C）は，二つの多目的ホール（0.71 s と 1.42 s）とクラッシックコンサートホール（1.96 s）の残響インパルス応答に対する結果を示す。

図11は，図10に示した三つの室内残響インパルス応答とそれらのパワーエンベロープを示す。結果として，残響インパルス応答のパワーエンベロープの近似度合が，残響時間 T_R の正確な推定に影響を与えていることがわかった。この結果は，特にラベルBで顕著に見られた。これは，提案法が1次近似として実環境で T_R をブラインド推定可能であることを示しているが，Bの場合，残響インパルス応答のパワーエンベロープ（式(7)）がもっと正確にモデル化されなければいけない。

8. 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3稿として，基本周波数推定法と残響時間推定法を概説し，それぞれの研究課題において現状でどの程度のことまでが達成されたのか，また残された課題が何であるかを説明した。

全3回を通じ，Houtgast と Steeneken によって導入された MTF の考えから，MTF に基づいた様々な音声信号処理の原理を概説してきた。このアプローチは決して万能なものであるわけではないが，音声明瞭度の観点から，室など環境が信号に与える影響を MTF としてモデル化することで，非常によく解けるケースがあることが魅力である。この良いケースと悪いケースの決定的な違いは，結局のところ室の伝達特性を MTF としてどのように関数化できるかであり，その逆特性もまたシンプルであることが望まれる。全3稿には，現状の課題等も含め概説していることから，これらを踏まえ，現手法の発展や改良法，あるいは新しい研究のシーズが生まれることを期待したい。

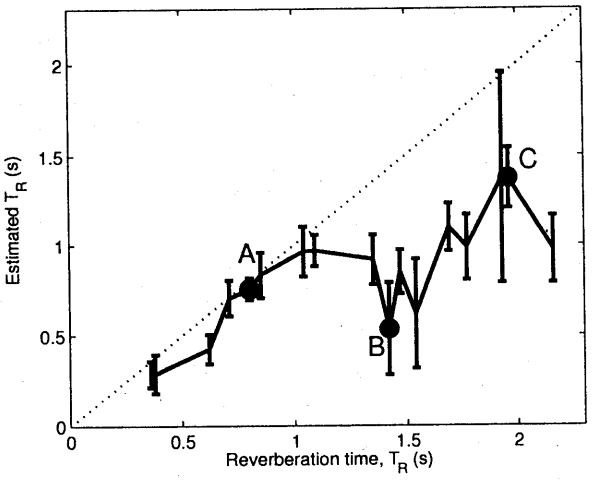


図 10　実際の残響環境における残響時間推定。A, B, C 点はそれぞれ多目的ホール（$T_{60} = 0.80$ s），多目的ホール（$T_{60} = 1.42$ s），教室（$T_{60} = 1.96$ s）の残響インパルス応答に対する結果である

Fig. 10 Estimated reverberation time in real reverberant environments. A, B, and C correspond to results for IRs in multipurpose halls (0.80 s), multipurpose halls (1.42 s), and classic concert halls (1.96 s)

謝　辞

本号で紹介した研究は，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（現在，ATR-SLT 勤務），Vu tat Thag 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，衣笠光太君，森田翔太君に心より感謝する。

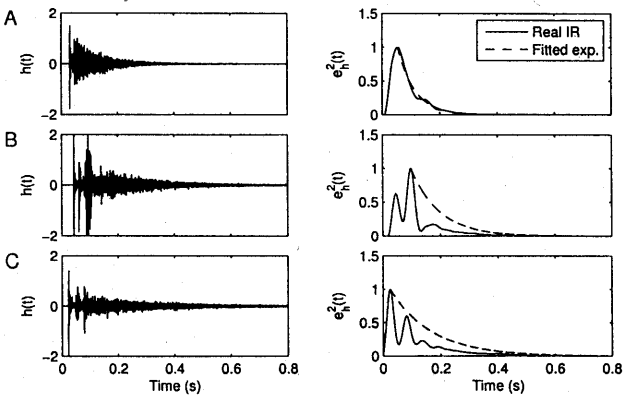


図 11　図 10 で示された各点の残響インパルス応答とそのパワーエンベロープ

Fig. 11 Power envelopes and impulse responses $h(t)$s at three points (A, B, and C) swhon in Fig. 10

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—, Journal of Signal Processing, Vol. 12, No. 5, pp. 339–348, 2008.

[2] 鵜木祐史: 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—, Journal of Signal Processing, Vol. 13, No. 1, pp. 3–12, 2009.

[3] W. J. Hess: Pitch Determination of Speech Signals, Springer-Verlag, New York, 1983.

[4] W. J. Hess: Pitch and Voicing Determination, in Advances in speech signal processing, Eds. S. Furui and M. M. Sondhi, pp. 3–48, Marcel Dekker. Inc. New York, 1992.

[5] A. de Cheveigné and H. Kawahara: Comparative evaluation of F0 estimation algorithms, Proc. Eurospeech2001, 2451–2454, 2001.

[6] A. M. Noll: Ceptrum pitch determination, J. Acoust. Soc. Am., Vol. 41, No. 2, pp. 293–309, 1966.

[7] T. V. Ananthapadmanabha and B. Yegnanarayana: Epoch extraction from linear prediction residual for indentification of closed glottis interval, IEEE Trans. Acoustics, Speech, Signal Processing, Vol. ASSP-27, No. 4, pp. 309–319, 1979.

[8] H. Kawahara, H. Katayose, A. de Cheveigné and R. D. Patterson: Fixed Point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity, Proc. Eurospeech99, Vol. 6, pp. 2781–2784, 1999.

[9] A. de Cheveigné and H. Kawahara: Yin, a fundamental frequency estimator for speech and music, J. Acoust. Soc. Am., Vol. 111, No. 4, pp. 1917–1930, 2002.

[10] Y. Atake, T. Irino, H. Kawahara, J. Lu, S. Nakamura and K. Shikano: Robust fundamental frequency estimation using instantaneous frequencies of harmonic components, Proc. ICSLP2000, Vol. 2, pp. 907–910, 2000.

[11] Y. Ishimoto, M. Unoki and M. Akagi: A fundamental frequency estimation method for noisy speech based on instantaneous amplitude and frequency, Proc. EuroSpeech2001, pp. 2439–2442, 2001.

[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] M. Unoki, T. Hosorogiya and Y. Ishimoto: Comparative evaluation of robust and accurate F0 estimates in reverberant environments, Proc. ICASSP2008, pp. 4569–4572, Las Vegas, Apr. 2008.

[14] H. Kuttruff: Room Acoustics, 3rd ed. (Elsevier Science Publishers Ltd., Lindin), 1991.

[15] ISO 3382, Acoustics–Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters, 2nd ed. (International Organization for Standardization, Géneve), 1997.

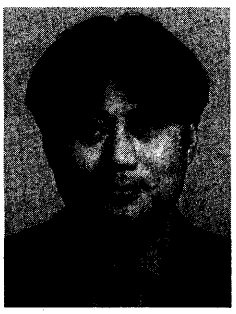
[16] J. Ohga, Y. Yamasaki and Y. Kaneda: Acoustic Systems and Digital Processing for Them, IEICE, Tokyo, 1995.

[17] S. Hiramatsu and M. Unoki: A study on the blind estimation of reverberation time in room acoustics, J. Signal Processing, Vol. 12, No. 6, pp. 351–361, 2008.

[18] M. Unoki and S. Hiramatsu: MTF-based method of blind estimation of reverberation time in room acoustics, Proc. EUSIPCO2008, Lausanne, Switzerland, Aug. 2008 (CD-ROM).

[19] T. Takeda, Y. Sagisaka, K. Katagiri, M. Abe and H. Kuwabara: Speech database user's manual, ATR Technical Report, TR-I-0028, 1988.

[20] SMILE2004, Sound Material in Living Environment, Architectural Institute of Japan and Gihodo Shuppan Co., Ltd., 2004.



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）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，音響電子透かしなど）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。