Title	残響音声からの基本周波数推定に関する検討
Author(s)	鵜木，祐史；石本，祐一；赤木，正人
Citation	Research report (School of Information Science, Japan Advanced Institute of Science and Technology), IS-RR-2005-007: 1-27
Issue Date	2005-03-28
Type	Technical Report
Text version	publisher
URL	http://hdl.handle.net/10119/8405
Rights	
Description	リサーチレポート（北陸先端科学技術大学院大学情報科学研究科）

残響音声からの基本周波数推定に関する検討

鵜木祐史，石本祐一，赤木正人

28 March 2005
IS-RR-2005-007

School of Information Science
Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, 923-1292, JAPAN
unoki@jaist.ac.jp, y-ishi@jaist.ac.jp, akagi@jaist.ac.jp

ISSN 0918-7553

残響音声からの基本周波数推定に関する検討

鵜木　祐史† 　石本　祐一† 　赤木　正人†

† 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
E-mail: †{unoki,y-ishi,akagi}@jaist.ac.jp

A study on an F0 estimation method for the reverberant speech

Masashi UNOKI†, Yuichi ISHIMOTO†, and Masato AKAGI†

† School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923–1292 Japan
E-mail: †{unoki,y-ishi,akagi}@jaist.ac.jp

1. はじめに

計算機における音声合成，音声認識，波形分離といった様々な音声信号処理技術において，音声の基本周波数は，音声信号の音源情報を担っていることから，非常に重要な特徴として利用されている．中でも，実環境において対象となる音声の基本周波数をクリーンな環境で抽出したものと同じ位，精度よく抽出できれば，実環境における様々な音声信号処理で生じる諸問題を解決するための重要な特徴として，推定された基本周波数を利用できる．例えば，実環境下の音声認識は，雑音や残響の影響により認識精度が著しく低下することが知られているが，基本周波数を句や単語の区分化，雑音除去・音声強調の手がかり，あるいは認識器に対するもう一つの特徴として利用することで，認識精度の向上に寄与することができる．そのため，様々な状況において，観測波形から基本周波数を正確に抽出することは，極めて重要な問題となっている [45].

基本周波数抽出（推定）に関する研究は古くから行われている研究課題であるが，現在でもまだ完全に解決されていないホットな話題である．この問題を複雑にしている原因は，大別すると (1) 音声が口唇から放射された音波であるため，基本周波数を表す声帯振動（音源情報）を直接観測することができず，しかもこの振動が準周期的であるということと，(2) 観測した音声信号には，雑音や残響の影響が混入しており，これらが音源情報を表す特徴を歪ませていることから正確な推定が困難であるということである．

これまでの研究は，主に (1) に主眼が置かれていたため，対象となる音声信号がクリーンな環境で観測されるものと暗黙の了解で仮定されていた．過去 50 年間，相当数の基本周波数抽出（推定）法が提案されてきた [1-4] が，最近ではこれらの研究の集大成として，実測された声帯振動（例えば，EGG）の情報と比較して，高信頼性を有し，高精度に基本周波数を推定できる方法が報告されている（例えば，TEMPO [33-36], YIN [8] など）．しかし，現実的な問題に対応するためには，(2) にも目を向けなければならならないし，雑音・残響環境下でこれらの高信頼性・高精度な基本周波数推定法がどれだけ正しく機能するのか見極めなければならない．

著者らは，これまでに様々な推定法に対する耐雑音性を調査し，いずれも低 SNR（音声に対し雑音パワーがかなり高い状況）では機能しないことを示した．また，これらの調査結果を踏まえ，耐雑音性に優れた基本周波数推定法を提案し，その有効性を示してきた [45]．耐雑音性を有する基本周波数推定法については，最近になって他にも報告されてきているが，残響に対する耐性についてはこれまでまったく議論されてこなかった．

本稿では，これまでに提案されてきた代表的な，高信頼で高精度な基本周波数推定法について，残響特性に対する耐性を調査する．次にこの調査結果に基づき，耐残響性に優れる特徴を洗い出し，高信頼性で高精度な基本周波数推定の方略を提案する．その後で，この方略に基づいたプロトタイプモデルを実装し，推定法を評価することでその有効性を示す．

2. 推定法の説明に必要な数式の定義

2.1 信号表現

観測される音声信号 $f(t)$ は，式 (1) のように，ソースフィルタモデルに基づき，その音源情報 $e(t)$ とある時刻 τ における声

道情報 $v_\tau(t)$ の畳み込みで表されるものとする．一方で，$f(t)$ を解析信号と見なし，式 (2) のように，振幅情報 $\alpha_k(t)$ と位相情報 $k\cdot\omega_0(t)$，φ_k からなる複素正弦波モデルとして表されるものとする．

$$
\begin{aligned}
f(t) &:= e(t) * v_\tau(t) &(1)\\
&:= \sum_k \alpha_k(t)\exp(jk\cdot\omega_0(t)t+\varphi_k) &(2)
\end{aligned}
$$

ここで，$\omega_0(t)$ は基本周波数 $F_0(t)$ に対する基本角周波数，φ_k は初期位相，k は信号の高調波成分数を表す．$f(t)$ は解析的な複素正弦波信号であるが，それぞれ，瞬時振幅，瞬時位相として個別に高調波信号を取り扱うことができる．このとき，解くべき問題は，観測信号 $f(t)$ から高調波を構成する基本波の周波数変調（あるいは位相変調）をもつ基本周波数 $F_0(t)$ を推定することである．

$$
F_0(t)=\frac{\omega_0(t)}{2\pi} \qquad (3)
$$

式 (2) に示す解析信号 $f(t)$ を，観測側で同じ形式で表現するために，一般に Fourier 変換対や短時間 Fourier 変換対，Wavelet 変換対等を利用する．$f(t)$ は，便宜上，時刻 τ での波形の切り出しとして窓関数 $w(t)$ を利用して解析される．ここで，式 (1) と式 (2) に対応する，切り出された被解析信号 $x(t)$ は，

$$
x(t,\tau)=w(t-\tau)f(t)
$$

で表されるものとする．また，音源信号 $s(\omega,\tau)$ と声道特性 $h(\omega,\tau)$ は，それぞれ

$$
\begin{aligned}
s(t,\tau) &= w(t-\tau)e(t)\\
h(t,\tau) &= v_\tau(t)
\end{aligned}
\qquad (4)
$$

で表されるものとする．

2.2 Fourier 変換／ Wavelet 変換によるスペクトル分析

式 (1) と式 (2) の信号表現に基づき，Fourier 変換対／ Wavelet 変換対を利用した信号の分析表現を示す．ここで，各変換の演算子を次のように定義する．

$\mathcal{F},\mathcal{F}^{-1}$: Fourier 変換, Fourier 逆変換

$\mathcal{S},\mathcal{S}^{-1}$: 短時間 Fourier 変換, 短時間 Fourier 逆変換

$\mathcal{W},\mathcal{W}^{-1}$: Wavelet 変換, Wavelet 逆変換

2.2.1 Fourier 変換対

$x(t,\tau)$ に対する Fourier 変換対は次式で表される．

$$
\begin{aligned}
X(\omega,\tau) &:= \mathcal{F}[x(t,\tau)]\\
&= \frac{1}{\sqrt{2\pi}}\int x(t,\tau)e^{-j\omega t}dt &(5)\\
x(t,\tau) &= \mathcal{F}^{-1}[X(\omega,\tau)]\\
&= \frac{1}{\sqrt{2\pi}}\int X(\omega,\tau)e^{j\omega t}d\omega &(6)
\end{aligned}
$$

ここで，$X(\omega,\tau)$ は複素スペクトルである．

2.2.2 短時間 Fourier 変換対

解析信号 $f(t)$ の短時間 Fourier 変換対は次式で表される．

$$
\begin{aligned}
X(\omega,\tau) &:= \mathcal{S}[f(t)]\\
&= \frac{1}{\sqrt{2\pi}}\int x(t)w(t-\tau)e^{-j\omega t}dt &(7)\\
f(t) &= \mathcal{S}^{-1}[X(\omega,\tau)]\\
&= \frac{1}{\sqrt{2\pi}}\int\int X(\omega,\tau)w(t-\tau)e^{j\omega t}d\omega d\tau &(8)
\end{aligned}
$$

2.2.3 Wavelet 変換対

解析信号 $f(t)$ の Wavelet 変換対は次式で表される．

$$
\begin{aligned}
F(a,b) &:= \mathcal{W}[f(t)]\\
&= \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(t)\overline{\psi\left(\frac{t-b}{a}\right)}dt &(9)\\
f(t) &= \mathcal{W}^{-1}[F(a,b)]\\
&= \frac{1}{D_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(a,b)\psi\left(\frac{t-b}{a}\right)\frac{dadb}{a^2} &(10)
\end{aligned}
$$

ここで，$F(a,b)$ は複素スペクトルであり，a はスケール，b はダイレーション（シフト）を表すパラメータである．また，$\psi(t)$ は基本 wavelet（あるいは analyzing wavelet）と呼ばれ，wavelet 変換の基底を表すものである．基本 wavelet としては任意の関数を設定することができるが，基底を成立させるために，アドミッシブル条件：

$$
D_\psi := \int_{-\infty}^{\infty}\frac{|\hat{\psi}(\omega)|}{|\omega|}d\omega<\infty \qquad (11)
$$

を満たす必要がある．これは，基底関数が時間的に閉じていて，その平均値が 0 であればよいことを意味している．

2.2.4 振幅特性／位相特性の表現

上記で示した複素スペクトルは，一般に実部と虚部の表現方法によって振幅特性と位相特性に分けて表現することもできる．例えば，式 (6) の複素スペクトル $x(\omega,\tau)$ を

$$
\begin{aligned}
X(\omega,\tau) &= |X(\omega,\tau)|\exp(j\arg X(\omega,\tau))\\
&= A(\omega,\tau)\exp(j\phi(\omega,\tau)) &(12)
\end{aligned}
$$

と表せば，その振幅スペクトル $A(\omega,\tau)$ と位相スペクトル $\phi(\omega,\tau)$ を

$$
\begin{aligned}
A(\omega,\tau) &= |X(\omega,\tau)| &(13)\\
\phi(\omega,\tau) &= \arctan\left(\frac{\Im[X(\omega,\tau)]}{\Re[X(\omega,\tau)]}\right) &(14)
\end{aligned}
$$

で求めることができる．ただし，$\Re\{\cdot\}$ と $\Im\{\cdot\}$ はそれぞれ，実部と虚部の成分を取り出す演算子である．尚，wavelet 変換対における複素スペクトル $F(a,b)$ も上記と同様に表現することができる．

2.3 複素ケプストラム分析

次に，2.2 節で述べたスペクトル分析について，(複素）ケプストラム分析により，別の情報表現をみてみる．

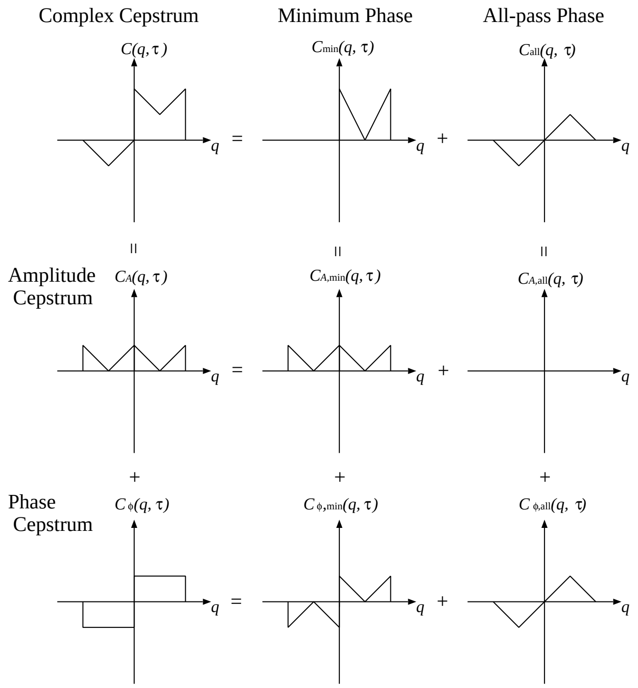

図 1　複素ケプストラム分析における振幅／位相特性と最小位相／全域通過特性の関係.

2. 3. 1　振幅スペクトルと位相スペクトルの分離

複素ケプストラム $C(q,\tau)$ は，対数複素スペクトル $\log X(\omega,\tau)$ に対する Fourier 逆変換として定義される.

$$C(q,\tau) := \mathcal{F}^{-1}\Big[\log X(\omega,\tau)\Big] \tag{15}$$

ここで，ケプストラム分析では，スペクトル分析での名前の逆を取る習慣から，パラメータはケフレンシー q（単位は時間）である．この複素ケプストラムを振幅スペクトル／位相スペクトルを利用して記述しなおすと

$$\begin{aligned} C(q,\tau) &= \mathcal{F}^{-1}\left[\log\left\{|X(\omega,\tau)|\exp(j\phi(\omega,\tau))\right\}\right] \\ &= \mathcal{F}^{-1}\Big[\log|X(\omega,\tau)|\Big] + \mathcal{F}^{-1}\Big[j\phi(\omega,\tau)\Big] \\ &= \mathcal{F}^{-1}\Big[\log A(\omega,\tau)\Big] + \mathcal{F}^{-1}\Big[j\phi(\omega,\tau)\Big] \end{aligned} \tag{16}$$

となる．この関係式を

$$C(q,\tau) = C_A(q,\tau) + C_\phi(q,\tau) \tag{17}$$

$$C_A(q,\tau) = \mathcal{F}^{-1}\Big[\log A(\omega,\tau)\Big] \tag{18}$$

$$C_\phi(q,\tau) = \mathcal{F}^{-1}\Big[j\phi(\omega,\tau)\Big] \tag{19}$$

と表すことにする．ここで，$C_A(q,\tau)$ は振幅ケプストラム，$C_\phi(q,\tau)$ は位相ケプストラムであり，各ケプストラムは図 1 の左側一列に示すような特徴をもつ．複素スペクトルでは振幅スペクトルと位相スペクトルの積で表現されていたのに対し，複素ケプストラムでは，振幅ケプストラムと位相ケプストラムの和で表現されることになる．この結果から複素ケプストラムを利用すれば，振幅特性と位相の関係を，スペクトルで積の表現からケプストラムで和の表現に変えて，信号を分析することができる．

更に，複素ケプストラムから複素スペクトル表現に戻すとき

$x(t,\tau)$	$=$	$x_{\min}(t,\tau)$	$*$	$x_{\rm all}(t,\tau)$
(Periodic)		(Minimum-Phase Component)		(All-Pass Component)

(Time-domain)

$\Downarrow$ $\mathcal{F}$ $\Uparrow$ $\mathcal{F}^{-1}$

$X(\omega,\tau)$	$=$	$X_{\min}(\omega,\tau)$	$\times$	$X_{\rm all}(\omega,\tau)$
(Complex)		(Complex)		(Complex)
$\vert\vert$		$\vert\vert$		$\vert\vert$
$\vert X(\omega,\tau)\vert$	$=$	$\vert X_{\min}(\omega,\tau)\vert$	$\times$	$\vert X_{\rm all}(\omega,\tau)\vert$
(Real)		(Real)		(Real)
$\times$		$\times$		$\times$
$e^{j\phi(\omega,\tau)}$	$=$	$e^{j\phi_{\min}(\omega,\tau)}$	$\times$	$e^{j\phi_{\rm all}(\omega,\tau)}$
(Complex)		(Complex)		(Complex)

(Frequency domain)

$\Downarrow$ log $\Uparrow$ exp

$\log X(\omega,\tau)$	$=$	$\log X_{\min}(\omega,\tau)$	$+$	$\log X_{\rm all}(\omega,\tau)$
(Complex)		(Complex)		(Complex)
$\vert\vert$		$\vert\vert$		$\vert\vert$
$\log\vert X(\omega,\tau)\vert$	$=$	$\log\vert X_{\min}(\omega,\tau)\vert$	$+$	$\log\vert X_{\rm all}(\omega,\tau)\vert$
(Real)		(Real)		(Real)
$+$		$+$		$+$
$j\phi(\omega,\tau)$	$=$	$j\phi_{\min}(\omega,\tau)$	$+$	$j\phi_{\rm all}(\omega,\tau)$
(Imagenal)		(Imagenal)		(Imagenal)

(Log-frequency domain)

$\Downarrow$ $\mathcal{F}^{-1}$ $\Uparrow$ $\mathcal{F}$

$C(\omega,\tau)$	$=$	$C_{\min}(\omega,\tau)$	$+$	$C_{\rm all}(\omega,\tau)$
(Asymmetric)		(Asymmetric)		(Asymmetric)
$\vert\vert$		$\vert\vert$		$\vert\vert$
$C_A(\omega,\tau)\vert$	$=$	$C_{A,\min}(\omega,\tau)\vert$	$+$	$C_{A,\rm all}(\omega,\tau)\vert$
(Even func.)		(Even func.)		(Even func.)
$+$		$+$		$+$
$C_\phi(\omega,\tau)$	$=$	$C_{\phi,\min}(\omega,\tau)$	$+$	$C_{\phi,\rm all}(\omega,\tau)$
(Odd func.)		(Odd func.)		(Odd func.)

(Quefrency (time) domain)

図 2　複素スペクトルと複素ケプストラムの対応関係.

には，下記のような表現もある．

$$\begin{aligned} \log X(\omega,\tau) &= \log A(\omega,\tau) + j\phi(\omega,\tau) \\ \log A(\omega,\tau) &= \Re\Big\{\mathcal{F}\Big[C_A(q,\tau)\Big]\Big\} = \Re\Big\{\mathcal{F}\Big[C(q,\tau)\Big]\Big\} \\ \phi(\omega,\tau) &= \Im\Big\{\mathcal{F}\Big[C_\phi(q,\tau)\Big]\Big\} = \Im\Big\{\mathcal{F}\Big[C(q,\tau)\Big]\Big\} \end{aligned}$$

ただし，$\Re\{\cdot\}$ と $\Im\{\cdot\}$ はそれぞれ，実部と虚部の成分を取り出す演算子である．また，複素ケプストラムと複素スペクトルの間には，Hilbert 変換（奇関数を偶関数に，偶関数を奇関数に変換）で結ばれた関係があり，

$$\Re[\cdot] = \text{Hilbert}\left[\Im[\cdot]\right] \tag{20}$$

$$\Im[\cdot] = \text{Hilbert}\left[\Re[\cdot]\right] \tag{21}$$

と表すことができる．これは，図 1 左側一列に見られるよう

に，振幅ケプストラムは偶関数，位相ケプストラムは奇関数となり，図 2 の左側一列に示すような計算過程から得られる．式 (20) と式 (21) は，それぞれ，図 2 の周波数領域の関係図とケフレンシー領域の関係図示される各記号の下の（ ）内の特徴の対応関係を表している．

2. 3. 2 最小位相特性と全域通過特性の分離

複素スペクトル表現において，振幅スペクトルと位相スペクトルの分割表現の他に，最小位相特性／全域通過特性による分割表現がある．$X_{\min}(\omega,\tau)$ を最小位相（複素）スペクトル，$X_{\mathrm{all}}(\omega,\tau)$ を全域通過（複素）スペクトルとすると，複素スペクトルは

$$\begin{aligned} X(\omega,\tau) &= X_{\min}(\omega,\tau)\cdot X_{\mathrm{all}}(\omega,\tau) \\ \log X(\omega,\tau) &= \log X_{\min}(\omega,\tau)+\log X_{\mathrm{all}}(\omega,\tau) \end{aligned}$$

と表される．ここで，上式を複素ケプストラムに対応づけると

$$\begin{aligned} C(q,\tau) &= \mathcal{F}^{-1}\left[\log X(\omega,\tau)\right] \\ &= \mathcal{F}^{-1}\left[\log X_{\min}(\omega,\tau)\right]+\mathcal{F}^{-1}\left[\log X_{\mathrm{all}}(\omega,\tau)\right] \end{aligned}$$

を得ることになり，これをまとめると

$$C(q,\tau) = C_{\min}(q,\tau)+C_{\mathrm{all}}(q,\tau) \tag{22}$$

$$C_{\min}(q,\tau) = \mathcal{F}^{-1}\left[\log X_{\min}(\omega,\tau)\right] \tag{23}$$

$$C_{\mathrm{all}}(q,\tau) = \mathcal{F}^{-1}\left[\log X_{\mathrm{all}}(\omega,\tau)\right] \tag{24}$$

と表すことができる．ここで，$C_{\min}(q,\tau)$ は最小位相（複素）ケプストラム，$C_{\mathrm{all}}(q,\tau)$ は全域通過位相（複素）ケプストラムである．このときのケプストラムの構成関係は図 1 の上段一行のようになり，複素スペクトルとの導出過程は，図 2 の上段一行となる．複素ケプストラムの偶関数／奇関数の関係から，最小位相（複素）ケプストラムは，

$$C_{\min}(q,\tau)=\begin{cases} 2C_A(q,\tau), & q>0 \\ C_A(q,\tau), & q=0 \\ 0, & \text{otherwise} \end{cases} \tag{25}$$

より容易に導出することができる．この関係から，最小位相（複素）スペクトルと全域通過（複素）スペクトルの振幅スペクトルと位相スペクトルは，それぞれ

$$\begin{aligned} \log|X_{\min}(\omega,\tau)| &= \Re\left\{\mathcal{F}\left[C_{\min}(q,\tau)\right]\right\} \\ \phi_{\min}(\omega,\tau) &= \Im\left\{\mathcal{F}\left[C_{\min}(q,\tau)\right]\right\} \\ \log|X_{\mathrm{all}}(\omega,\tau)| &= \Re\left\{\mathcal{F}\left[C_{\mathrm{all}}(q,\tau)\right]\right\} \\ \phi_{\mathrm{all}}(\omega,\tau) &= \Im\left\{\mathcal{F}\left[C_{\mathrm{all}}(q,\tau)\right]\right\} \end{aligned}$$

となる．これらの関係を図 1 の中央一列と右側一列に示す．

2. 3. 3 最小位相成分と全域通過位相成分の分離

上記で得られた最小位相（複素）ケプストラムと全域通過位相（複素）ケプストラムから，一旦，該当する複素スペクトルに戻して表現することにする．ここで，最小位相（複素）スペクトルを $X_{\min}(\omega,\tau)$，全域通過（複素）スペクトルを $X_{\mathrm{all}}(\omega,\tau)$ とすると，

$$\begin{aligned} X_{\min}(\omega,\tau) &= |X_{\min}(\omega,\tau)|\exp(j\arg X_{\min}(\omega,\tau)) \\ &= A_{\min}(\omega,\tau)\exp(j\phi_{\min}(\omega,\tau)) \end{aligned} \tag{26}$$

$$\begin{aligned} X_{\mathrm{all}}(\omega,\tau) &= |X_{\mathrm{all}}(\omega,\tau)|\exp(j\arg X_{\mathrm{all}}(\omega,\tau)) \\ &= A_{\mathrm{all}}(\omega,\tau)\exp(j\phi_{\mathrm{all}}(\omega,\tau)) \end{aligned} \tag{27}$$

と表すことができる．このとき，全域通過特性をもつ $X_{\mathrm{All}}(\omega,\tau)$ は大きさが 1 （$A_{\mathrm{all}}(\omega,\tau)=|X_{\mathrm{all}}(\omega,\tau)|=1$）であるため，複素スペクトルの大きさ $A(\omega,\tau)$ は最小位相（複素）スペクトルの大きさと等しいことがわかる．

$$A(\omega,\tau)=A_{\min}(\omega,\tau)\cdot A_{\mathrm{all}}(\omega,\tau)=A_{\min}(\omega,\tau) \tag{28}$$

この結果に従い，再度，複素スペクトルを最小位相／全域通過位相に関して，振幅スペクトルと位相スペクトルで表現すると，

$$\begin{aligned} X(\omega,\tau) &= X_{\min}(\omega,\tau)\cdot X_{\mathrm{all}}(\omega,\tau) \\ &= |X_{\min}(\omega,\tau)|\exp(j\phi_{\min}(\omega,\tau)) \\ &\qquad \times|X_{\mathrm{all}}(\omega,\tau)|\exp(j\phi_{\mathrm{all}}(\omega,\tau)) \\ &= A(\omega,\tau)\exp(j\phi_{\min}(\omega,\tau))\cdot\exp(j\phi_{\mathrm{all}}(\omega,\tau)) \end{aligned}$$

となるため，最終的に，

$$\log X(\omega,\tau)=\log A(\omega,\tau)+j\phi_{\min}(\omega,\tau)+j\phi_{\mathrm{all}}(\omega,\tau) \tag{29}$$

となる．また，同じ表現を複素ケプストラム表現上ですると

$$\begin{aligned} C(q,\tau) &= \mathcal{F}^{-1}\Big[\log X(\omega,\tau)\Big] \\ &= \mathcal{F}^{-1}\Big[\log|X_{\min}(\omega,\tau)|\Big]+\mathcal{F}^{-1}\Big[j\phi_{\min}(\omega,\tau)\Big] \\ &\quad +\mathcal{F}^{-1}\Big[\log|X_{\mathrm{all}}(\omega,\tau)|\Big]+\mathcal{F}^{-1}\Big[j\phi_{\mathrm{all}}(\omega,\tau)\Big] \\ &= \mathcal{F}^{-1}\Big[\log|X_{\min}(\omega,\tau)|\Big]+\mathcal{F}^{-1}\Big[j\phi_{\min}(\omega,\tau)\Big] \\ &\quad +\mathcal{F}^{-1}\Big[j\phi_{\mathrm{all}}(\omega,\tau)\Big] \end{aligned}$$

の導出過程から

$$\begin{aligned} C(q,\tau) &= C_{A,\min}(q,\tau)+C_{\phi,\min}(q,\tau) \\ &\qquad +C_{A,\mathrm{all}}(q,\tau)+C_{\phi,\mathrm{all}}(q,\tau) \\ &= C_A(q,\tau)+C_{\phi,\min}(q,\tau)+C_{\phi,\mathrm{all}}(q,\tau) && (30) \\ &= C_{\min}(q,\tau)+C_{\phi,\mathrm{all}}(q,\tau) && (31) \end{aligned}$$

を得る．ただし，

$$C_{\min}(q,\tau) = C_{A,\min}(q,\tau)+C_{\phi,\min}(q,\tau) \tag{32}$$

$$C_{\mathrm{all}}(q,\tau) = C_{\phi,\mathrm{all}}(q,\tau) \tag{33}$$

である．

以上，複素スペクトルと複素ケプストラムの関係を，最小位相特性と全域通過特性に着目し，更に振幅特性と位相特性に分離した場合の表現をまとめた．図 1 と図 2 の全般を眺めると，複素スペクトルの導出過程と複素ケプストラムの導出関係を通して，お互いの対応関係を容易に読み取ることができる．

2. 4 瞬時周波数

瞬時周波数 $f_i(\omega,\tau)$ は，次式で定義されるように，位相スペクトル $\phi(\omega,\tau)$ の時間微分で表される．

表 1　論文中で利用される記号の定義.

記号/変数	数学的な意味
t	時間 (s)
f	周波数 (Hz)
q	ケフレンシー (s)
$f(t)$	解析的な周期信号（有声成分）
$e(t)$	音源信号（励振信号）
$v_\tau(t)$	フィルタ特性（時不変インパルス応答）
$F_0(t)$	基本周波数 (Hz)
$T_0(t)$	基本周期 (s)
$\alpha_k(t)$	瞬時振幅
$\omega_0(t)$	基本角周波数
$w(t)$	窓関数
$\psi(t)$	基本 wavelet
$x(t,\tau)$	時刻 τ で切り出された解析信号
$s(t,\tau)$	時刻 τ での短時間音源信号
$h(t,\tau)$	時刻 τ でのフィルタのインパルス応答
$X(\omega,\tau)$	複素スペクトル
$A(\omega,\tau)$	振幅スペクトル
$\phi(\omega,\tau)$	位相スペクトル
$X_{\mathrm{min}}(\omega,\tau)$	最小位相特設をもつ複素スペクトル
$X_{\mathrm{all}}(\omega,\tau)$	全域通過特性をもつ複素スペクトル
$A_{\mathrm{min}}(\omega,\tau)$	最小位相特設をもつ振幅スペクトル
$A_{\mathrm{all}}(\omega,\tau)$	全域通過特性をもつ振幅スペクトル
$\phi_{\mathrm{min}}(\omega,\tau)$	最小位相特性をもつ位相スペクトル
$\phi_{\mathrm{all}}(\omega,\tau)$	全域通過特性をもつ位相スペクトル
$S(\omega,\tau)$	音源信号の複素スペクトル
$H(\omega,\tau)$	フィルタの伝達関数を表す複素スペクトル
$C(q,\tau)$	複素ケプストラム
$C_A(q,\tau)$	振幅ケプストラム
$C_\phi(q,\tau)$	位相ケプストラム
$C_{\mathrm{min}}(q,\tau)$	最小位相特性をもつ複素ケプストラム
$C_{\mathrm{all}}(q,\tau)$	全域通過特性をもつ複素ケプストラム
$C_{A,\mathrm{min}}(q,\tau)$	最小位相特性をもつ振幅ケプストラム
$C_{A,\mathrm{all}}(q,\tau)$	全域通過特性をもつ振幅ケプストラム
$C_{\phi,\mathrm{min}}(q,\tau)$	最小位相特性をもつ位相ケプストラム
$C_{\phi,\mathrm{all}}(q,\tau)$	全域通過特性をもつ位相ケプストラム
$C_{\mathrm{src}}(q,\tau)$	音源信号の複素ケプストラム
$C_{\mathrm{flt}}(q,\tau)$	フィルタ特性を表す複素ケプストラム
$f_i(\omega,\tau)$	瞬時周波数
$\theta(\omega,\tau)$	群遅延
$\theta_{\mathrm{min}}(\omega,\tau)$	最小位相成分における群遅延
$\theta_{\mathrm{all}}(\omega,\tau)$	全域通過成分における群遅延

$$f_i(\omega,\tau) := \frac{1}{2\pi}\frac{\partial \arg X(\omega,\tau)}{\partial \tau} \tag{34}$$
$$= \frac{1}{2\pi}\frac{\partial \phi(\omega,\tau)}{\partial \tau} \tag{35}$$

ただし，位相スペクトルは時間 τ に対して連続であるものとしている．

ここで，位相スペクトルと位相ケプストラムの対応関係から，ケプストラムも時間 τ について連続であると仮定できれば，

$$\phi(\omega,\tau) = \phi_{\mathrm{min}}(\omega,\tau) + \phi_{\mathrm{all}}(\omega,\tau)$$
$$\Updownarrow$$
$$C_\phi(q,\tau) = C_{\phi,\mathrm{min}}(q,\tau) + C_{\phi,\mathrm{all}}(q,\tau)$$

の関係から，位相スペクトルと位相ケプストラムの Hilbert 変換対は，

$$\begin{cases} C_\phi(q,\tau) & = \mathcal{F}^{-1}\left[j\phi(\omega,\tau)\right] \\ C_{\phi,\mathrm{min}}(q,\tau) & = \mathcal{F}^{-1}\left[j\phi_{\mathrm{min}}(\omega,\tau)\right] \\ C_{\phi,\mathrm{all}}(q,\tau) & = \mathcal{F}^{-1}\left[j\phi_{\mathrm{all}}(\omega,\tau)\right] \end{cases}$$
$$\Updownarrow$$
$$\begin{cases} \phi(\omega,\tau) & = \Im\left\{\mathcal{F}\left[C_\phi(q,\tau)\right]\right\} \\ \phi_{\mathrm{min}}(\omega,\tau) & = \Im\left\{\mathcal{F}\left[C_{\phi,\mathrm{min}}(q,\tau)\right]\right\} \\ \phi_{\mathrm{all}}(\omega,\tau) & = \Im\left\{\mathcal{F}\left[C_{\phi,\mathrm{all}}(q,\tau)\right]\right\} \end{cases}$$

で表される．つまり，この関係を瞬時周波数 $f_i(\omega,\tau)$ の定義式に代入すると，瞬時周波数 $f_i(\omega,\tau)$ をそれぞれ最小位相成分に対応する瞬時周波数 $f_{i,\mathrm{min}}(\omega,\tau)$ と全域通過位相成分に対応する瞬時周波数 $f_{i,\mathrm{all}}(\omega,\tau)$ で表すことができる．

$$\begin{aligned} f_i(\omega,\tau) &= \frac{1}{2\pi}\frac{\partial\phi(\omega,\tau)}{\partial\tau} \\ &= \frac{1}{2\pi}\frac{\partial}{\partial\tau}\Big\{\phi_{\mathrm{min}}(\omega,\tau) + \phi_{\mathrm{all}}(\omega,\tau)\Big\} \\ &= f_{i,\mathrm{min}}(\omega,\tau) + f_{i,\mathrm{all}}(\omega,\tau) \end{aligned}$$

ただし，

$$f_i(\omega,\tau) = \frac{1}{2\pi}\frac{\partial}{\partial\tau}\Big(\Im\Big\{\mathcal{F}\big[C(q,\tau)\big]\Big\}\Big) \tag{36}$$
$$f_{i,\mathrm{min}}(\omega,\tau) = \frac{1}{2\pi}\frac{\partial}{\partial\tau}\Big(\Im\Big\{\mathcal{F}\big[C_{\mathrm{min}}(q,\tau)\big]\Big\}\Big) \tag{37}$$
$$f_{i,\mathrm{all}}(\omega,\tau) = \frac{1}{2\pi}\frac{\partial}{\partial\tau}\Big(\Im\Big\{\mathcal{F}\big[C_{\mathrm{all}}(q,\tau)\big]\Big\}\Big) \tag{38}$$

である．

2.5 群　遅　延

位相特性の一つである群遅延 $\theta(\omega,\tau)$ は，次式に定義されるように，位相スペクトル $\phi(\omega,\tau)$ の周波数方向への微分で表される．

$$\theta(\omega,\tau) := -\frac{\partial \arg X(\omega,\tau)}{\partial\omega} \tag{39}$$
$$= -\frac{\partial\phi(\omega,\tau)}{\partial\omega} \tag{40}$$

ただし，位相スペクトルは ω に対して連続であるものとしている．

ここで，2.4 節と同様に，最小位相成分と全域通過位相成分に分離して表現すると

$$\begin{aligned} \theta(\omega,\tau) &= -\frac{\partial\phi(\omega,\tau)}{\partial\omega} \\ &= -\frac{\partial}{\partial\omega}\left\{\phi_{\mathrm{min}}(\omega,\tau) + \phi_{\mathrm{all}}(\omega,\tau)\right\} \\ &= \theta_{\mathrm{min}}(\omega,\tau) + \theta_{\mathrm{all}}(\omega,\tau) \end{aligned}$$

を得る．ただし，

$$\theta(\omega,\tau) = -\frac{\partial}{\partial\omega}\Big(\Im\Big\{\mathcal{F}\big[C(q,\tau)\big]\Big\}\Big) \tag{41}$$
$$\theta_{\mathrm{min}}(\omega,\tau) = -\frac{\partial}{\partial\omega}\Big(\Im\Big\{\mathcal{F}\big[C_{\mathrm{min}}(q,\tau)\big]\Big\}\Big) \tag{42}$$
$$\theta_{\mathrm{all}}(\omega,\tau) = -\frac{\partial}{\partial\omega}\Big(\Im\Big\{\mathcal{F}\big[C_{\mathrm{all}}(q,\tau)\big]\Big\}\Big) \tag{43}$$

である．

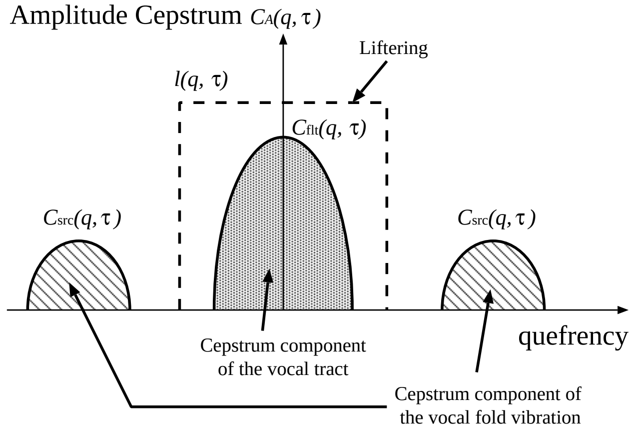

図 3　ケフレンシー領域におけるソース・フィルタモデルの分離.

2.6　ケプストラム分析を利用した音源／フィルタ特性の分離

2.3 節の複素ケプストラム分析で述べたように，複素スペクトルでは振幅スペクトルと位相スペクトルが積で表現されたのに対し，それぞれに対応する振幅ケプストラムと位相ケプストラムは和で表現された．最初の信号表現に戻り，式 (1) について再考すると，時間領域で畳み込みの関係にあるものは，スペクトルに変換した周波数領域では積の関係にある．つまり，音源情報 $e(t)$ と声道フィルタ特性 $v_\tau(t)$ の畳み込みは，$F(\omega) = E(\omega) \cdot V(\omega)$ と積で表現されることになる．これを短時間 Fourier 変換を利用して信号を分析したものとすると，

$$\begin{aligned} X(\omega,\tau) &= X_{\mathrm{src}}(\omega,\tau) \cdot X_{\mathrm{flt}}(\omega,\tau) \\ &= S(\omega,\tau) \cdot H(\omega,\tau) \end{aligned} \tag{44}$$

となる．ただし，音源情報を表すスペクトル $S(\omega,\tau)$ とフィルタ特性を表すスペクトル $H(\omega,\tau)$ は，それぞれ

$$\begin{aligned} S(\omega,\tau) &= X_{\mathrm{src}}(\omega,\tau) = |X_{\mathrm{src}}(\omega,\tau)| \exp(j\phi_{\mathrm{src}}(\omega,\tau)) \\ H(\omega,\tau) &= X_{\mathrm{flt}}(\omega,\tau) = |X_{\mathrm{flt}}(\omega,\tau)| \exp(j\phi_{\mathrm{flt}}(\omega,\tau)) \end{aligned}$$

である．そのため，時刻 τ での音源信号とインパルス応答は

$$s(t,\tau) = \mathcal{F}^{-1}\big[S(\omega,\tau)\big] \tag{45}$$

$$h(t,\tau) = \mathcal{F}^{-1}\big[H(\omega,\tau)\big] \tag{46}$$

と表される．

一方，複素ケプストラム分析から上式を捉え直すと

$$\begin{aligned} X(\omega,\tau) &= X_{\mathrm{src}}(\omega,\tau) \cdot X_{\mathrm{flt}}(\omega,\tau) \\ \log X(\omega,\tau) &= \log X_{\mathrm{src}}(\omega,\tau) + \log X_{\mathrm{flt}}(\omega,\tau) \\ &= \mathcal{F}\big[C_{\mathrm{src}}(q,\tau)\big] + \mathcal{F}\big[C_{\mathrm{flt}}(q,\tau)\big] \\ &= \mathcal{F}\big[C(q,\tau)\big] \end{aligned}$$

から，音源情報とフィルタ特性に対応する複素スペクトル $X_{\mathrm{src}}(\omega,\tau)$ と $X_{\mathrm{flt}}(\omega,\tau)$ は，スペクトルで積の表現を取るのに対し，それぞれに対応する複素ケプストラムは，ケプストラムで和の表現を取ることになる．

$$\begin{aligned} C(q,\tau) &:= C_{\mathrm{src}}(q,\tau) + C_{\mathrm{flt}}(q,\tau) \\ &= C_A(q,\tau) + C_\phi(q,\tau) \end{aligned} \tag{47}$$

ここで，式 (47) で表された複素ケプストラムの実部分の対応関係を図 3 に示す．対数振幅スペクトルにおける周波数方向への変動を表すものが実ケプストラムなのだから，一般的に，スペクトル上では，速い振幅変動を示す音源情報（調波性として，基本周波数 $F0$ に関する山谷の変化となって表れる）と緩やかな変動を示すフィルタの伝達特性（声道特性としてフォルマントやアンチフォルマントの配置にともなう変化）は，それぞれ高ケフレンシー領域と低ケフレンシー領域に分離して表れる．そのため，ケフレンシー領域では，音源かフィルタに対応するケプストラムを切り出すだけで，その特徴を取り出すことができる．このような切り出しに利用するフィルタは，リフターと呼ばれ，図 3 の破線ブロックで示される．

この考えを利用して，複素ケプストラムの特性を整理すると音源情報とフィルタ特性に対応する複素ケプストラム $C_{\mathrm{src}}(q,\tau)$ と $C_{\mathrm{flt}}(q,\tau)$ は

$$\begin{aligned} C_{\mathrm{src}}(q,\tau) &= C_{A,\mathrm{src}}(q,\tau) + C_{\phi,\mathrm{src}}(q,\tau) \\ C_{\mathrm{flt}}(q,\tau) &= C_{A,\mathrm{flt}}(q,\tau) + C_{\phi,\mathrm{flt}}(q,\tau) \\ C_A(q,\tau) &= C_{A,\mathrm{src}}(q,\tau) + C_{A,\mathrm{flt}}(q,\tau) \\ C_\phi(q,\tau) &= C_{\phi,\mathrm{src}}(q,\tau) + C_{\phi,\mathrm{flt}}(q,\tau) \end{aligned}$$

で表され，結果として

$$C_{A,\mathrm{src}}(q,\tau) := (1 - \ell(q,\tau))C_A(q,\tau) \tag{48}$$

$$C_{A,\mathrm{flt}}(q,\tau) := \ell(q,\tau)C_A(q,\tau) \tag{49}$$

$$C_{\phi,\mathrm{src}}(q,\tau) := C_{\phi,\mathrm{min}}(q,\tau) \tag{50}$$

$$C_{\phi,\mathrm{flt}}(q,\tau) := C_{\phi,\mathrm{all}}(q,\tau) \tag{51}$$

となる．ここで，$\ell(q,\tau)$ リフターは，音源情報とフィルタ特性を分離する部分にカットオフを持つように設計されるものとする．

例として，図 4(b),(c) に音源信号 $e(t)$ とフィルタのインパルス応答 $v(t)$ を示す．図 4(a) はこれらの畳み込み信号である．図 4(d)-(f) はそれぞれ，図 4(a)-(c) の振幅スペクトルを示す．この事例では，事前にスペクトル領域で作成し，それを時間領域に変換して表示してある．音源信号は基本周波数 100 Hz 一定，調波の最大次数を 50 次とし，フィルタ形状は 12 次の LPC を利用して母音/a/を模擬した．図 4(g)-(l) は，それぞれ上式に示した振幅ケプストラムと位相ケプストラムを示す．図 4(h) の 0.01 s に顕著なピークが見られるが，このときのケフレンシーがちょうど基本周期 T_0 に対応し，この逆数が基本周波数 F_0 に対応する．

この考えに基づくと，観測信号から声道フィルタ特性を除去した状態で音源情報のみから基本周波数を推定することが可能になる．このケプストラム分析の場合は，$C_{\mathrm{src}}(q,\tau)$ から Fourier 変換対を利用することで推定可能となる．

本稿では，LPC 分析，適応的フィルタリング等の数学的準備については割愛するが，上記のように声道フィルタ形状を模擬／推定し，その逆特性を利用することで音源情報のみを取り

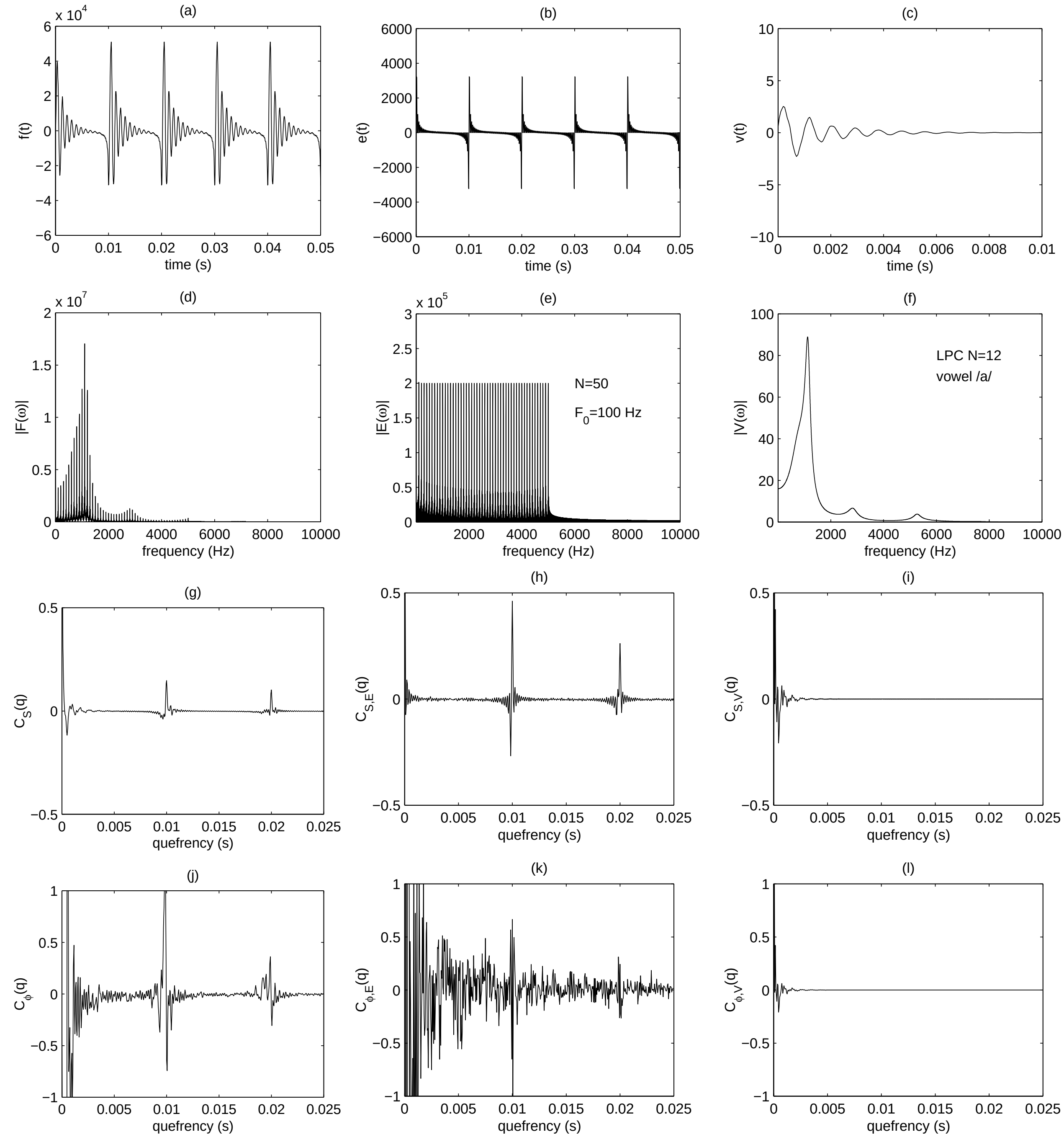

図 4　音源情報とフィルタ特性の関係．(a) 信号 $f(t)$, (b) 音源信号 $e(t)$, (c) フィルタのインパルス応答 $v(t)$, (d) 信号の振幅スペクトル $|F(\omega)|$, (e) 音源スペクトル $|E(\omega)|$, (f) フィルタの伝達関数 $|V(\omega)|$, (g) 信号の振幅ケプストラム $C_S(q)$, (h) 音源の振幅ケプストラム $C_{S,E}(q)$, (i) フィルタの振幅ケプストラム $C_{S,V}(q)$, (j) 信号の位相ケプストラム $C_\phi(q)$, (k) 音源の位相ケプストラム $C_{\phi,E}(q)$, (l) フィルタの位相ケプストラム $C_{\phi,V}(q)$.

出し，基本周波数を推定する方法が他にも知られている [1-4]. いずれも手法が異なるだけで，本質的な狙いは同じである．

3.　基本周波数推定法のアルゴリズム

ここでは，これまでに報告されてきた代表的な基本周波数推定法を紹介する．

3.1　波形処理における基本周波数推定法

3.1.1　時間波形から直接推定する方法

観測された音声信号そのものから直接，基本周波数を推定する方法として，図 5(a)-(c) に示すような，信号 $x(t,\tau)$ における (a) ゼロ交差法，(b) ピーク検出法，(c) 自己相関法がある．これらの手法はいずれも，窓関数 $w(t)$ で切り出した $x(t,\tau)$ における周期性の検出を狙ったものである．推定結果は基本周期 $T_0(t)$ となるため，この逆数を取ることで，基本周波数の推定

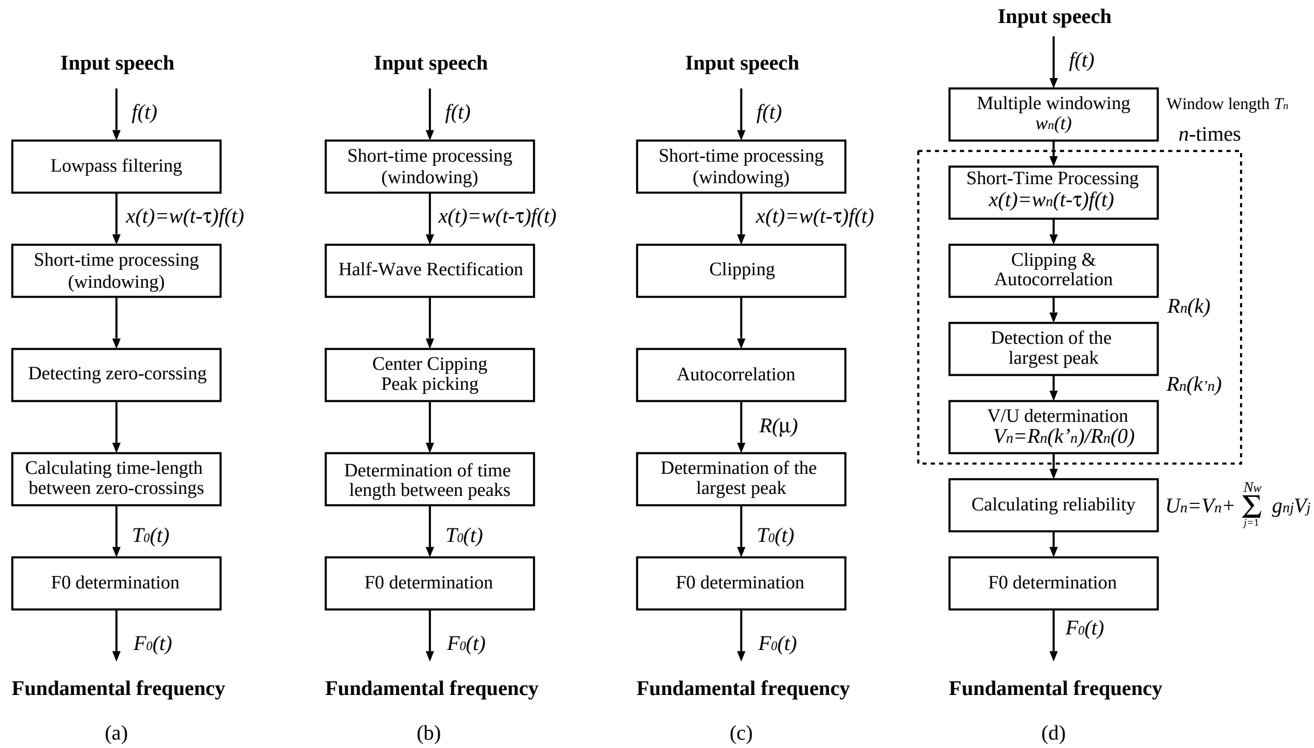

図 5　波形処理における基本周波数推定法．(a) ゼロ交差法，(b) ピーク検出法，(c) 自己相関法，(d) 複数窓長を利用した自己相関法のブロックダイアグラム．

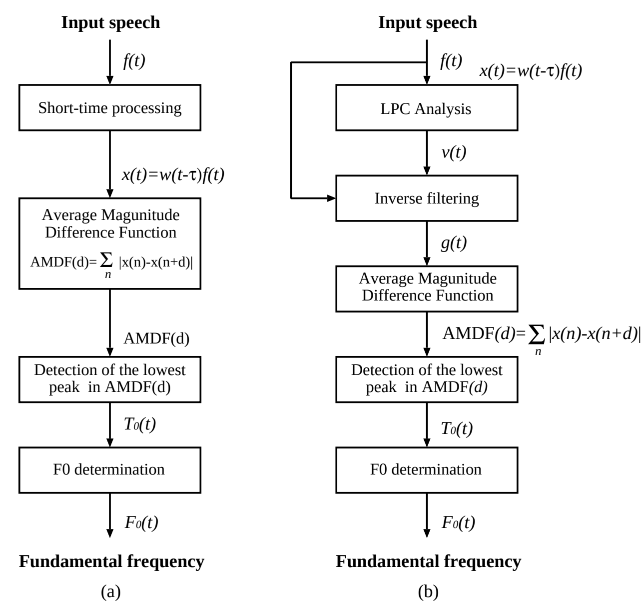

図 6　AMDF 法を利用した基本周波数推定法のブロックダイアグラム．(a) 時間波形に対する AMDF 法，(b) LPC 残差信号に対する AMDF 法．

値が得られる．

ゼロ交差法 [1, 2, 5] では，波形レベルでの振幅値 0 の交差点を微分処理等で検出し，その時間間隔 $T_0(t)$ を τ の間隔で検出するものである．ピーク検出法 [5, 6] は，同じく $x(t,\tau)$ の振幅値のピーク値を微分処理等で検出し，その時間間隔 $T_0(t)$ を τ の間隔で検出するものである．自己相関法 [1, 2, 8] は，次式に示す $x(t,\tau)$ の自己相関関数を利用してその相関値のピーク値を検出し，最大ピークを産み出すときの時刻 $\mu = \mu_0$ を求めることで，時間周期 $T_0(t) = \mu_0$ を得るものである．

$$R(\mu) = \sum_{t=t_0}^{t_0+T_\mu} x(t,\tau)x(t+\mu,\tau) \tag{52}$$

ただし，T_μ は自己相関関数を求める幅であり，一般に窓関数によって切り出した $x(t,\tau)$ の幅（つまり，窓長）になる．自己相関法には，変形自己相関法や対象となる $x(t,\tau)$ にクリッピング処理（例えば，$x(t,\tau)$ の時間平均値以下をゼロとするような非線形処理を施す）を利用するものなど，その改良は多枝に渡っている．

これらの処理には，窓関数による切り出しがあるが，一般にこの切り出し幅は基本周期より広くなければならず，通常 2，3 倍の幅を取ることが多いが男性話者と女性話者によってこの周期が 2 倍以上に異なることから窓長の選び方には注意を必要としている．

3.1.2　多重窓長を利用した自己相関法

窓長の設定によっては，得られるべき基本周期が得られなかったり，2 倍周期，半周期の基本周期を誤推定することがある．これを避けるために提案された手法が ACMWL (Auto-Correlation Multiple Window Length) 法 [9] である．この手法のブロックダイアグラムを図 5(d) に示す．図中の破線ブロック内は基本的に図 5(c) の自己相関法であり，数種類の選ばれた窓長を利用して，基本周波数を推定し，その中から適切な窓長と基本周波数を推定するものである．破線ブロック内は，図 5(a) や図 5(b) の手法を当てはめることも可能である．

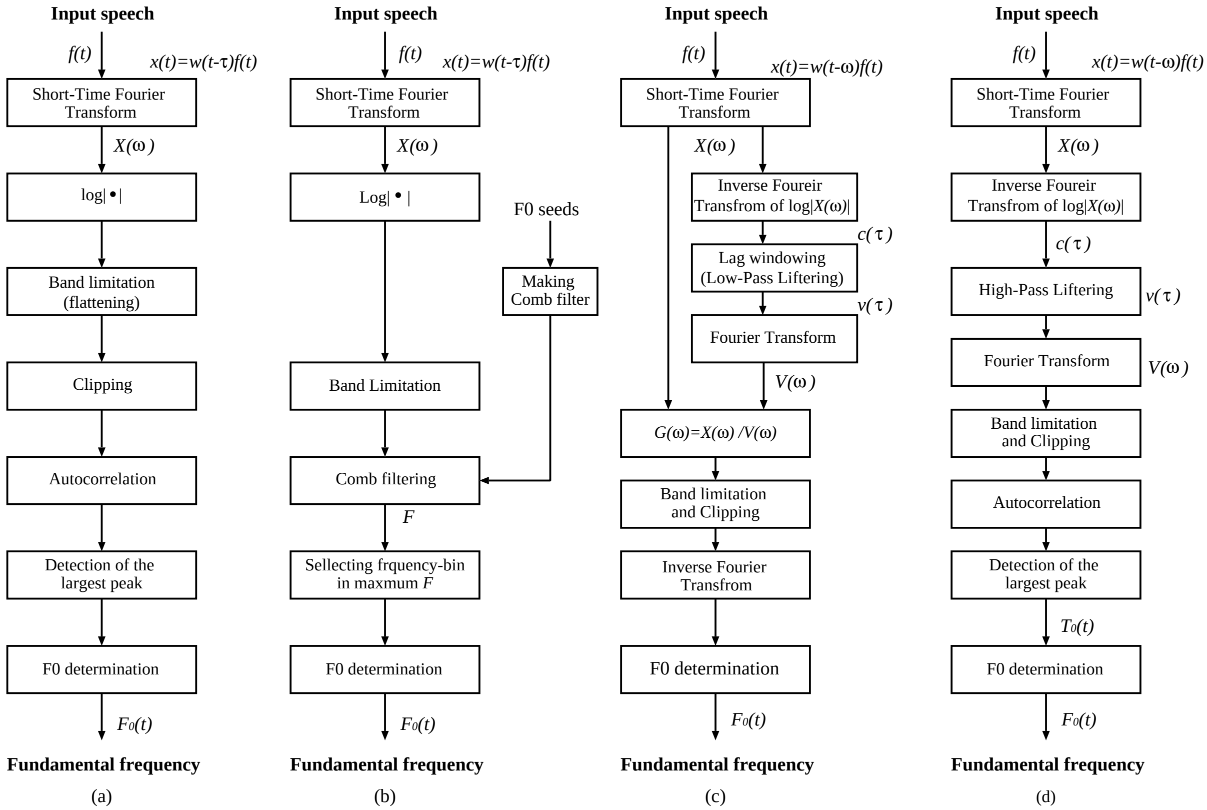

図 7　短時間 Fourier 変換を利用した基本周波数推定法のブロックダイアグラム．(a) 自己相関法，(b) Comb フィルタ法，(c) ラグ窓法，(d) リフター法.

3.1.3　AMDF 法

時間波形におけるもう一つの処理として，図 6 に示すような，平均振幅差関数 AMDF (Avarage Magnitude Difference Function) 法 [10] がある．これはある時刻 τ で窓関数で切り取られた $x(t,\tau)$ に関して，AMDF の距離尺度を利用して周期性を検出するものである．この他に，後述する LPC の残差信号に対しても同様に AMDF を利用する方法 [11] もある.

3.2　短時間 Fourier 変換を利用した推定方法

図 7 に，短時間 Fourier 変換（STFT）を利用した基本周波数推定法を示す．この方法は基本的に，2.2 節で説明した短時間 Fourier 変換により得られた振幅スペクトル $|X(\omega,\tau)|$ あるいは対数振幅スペクトル $\log|X(\omega,\tau)|$ に対し，音源信号の調波性を何らかの処理により求め，その基本波に対応する基本周波数 $F_0(t)$ を推定するものである．代表的な処理は，図 7(a) に示す自己相関法 [1-4, 19, 20] である．これは，対数振幅スペクトル $\log|X(\omega,\tau)|$ に対し，帯域制限，クリッピング処理（対数振幅スペクトルの周波数方向への平均値に対し，平均以下をゼロ埋めすることで非線形処理を施し，自己相関への寄与を強調する）をした上で，自己相関関数によりピークとなる周波数シフト位置を求めるものである．また，類似する手法として，図 7(b) に示すように，対数振幅スペクトル上の調波性を Comb フィルタ [22-25]（ある基本周波数を仮定してその調波にあたる櫛を作成するもの）を通し，その積和が最大になる基本周波数を推定値とするものもある [25].

一方，図 7(c),(d) に示すように，一度，対数振幅スペクトル $\log|X(\omega,\tau)|$ を求めた後，いわゆるケプストラム上でのラグ窓により声道特性を推定し，その逆特性をかけるか [21]，リフター処理（低域通過フィルタ）をかけることで声道特性の情報を打ち消し（対数振幅の白色化処理），復元された音源信号上あるいは，音源スペクトル上で，自己相関法などを利用し，その周期性あるいは調波性から基本周期，基本周波数を推定するものである [4]．ラグ窓関数を利用するものをラグ窓法（図 7(c)），リフター処理を利用するものをリフター法（図 7(d)）と呼ぶ.

3.3　ケプストラム処理を利用した推定方法

2.6 節で述べたケプストラム処理における音源／フィルタ特性の分離の考えに基づいた基本周波数推定法には，図 8 に示すように，(a) ケプストラム法（Noll の推定法 [12, 13]），(b) クリップストラム法 [14]，(c) 改良ケプストラム法 [1-4, 15] がある．これらは，振幅（実）ケプストラムのみに着目し，音源情報と声道特性を表すフィルタに分割し，その音源特性に見られる特徴から基本周期ならびに基本周波数を推定するものである.

この手法の最初のものは，Noll によって提案された処理体系であり，(a) の基本処理は，振幅ケプストラム $C_A(q,\tau)$ あるいは $C_{A,\mathrm{src}}(q,\tau)$ で観測される基本周期に対応するピーク位置を検出し，このときのケフレンシーから $T_0(t)$ を求め，結果として基本周波数 $F_0(t)$ を求めるものである．(b) のクリップストラムは，クリッピング処理とケプストラム処理の結合から，両者を文字って提案されたものであり，原理的にはケプストラム処理と同じである.

(c) の改良ケプストラム法には，(a) の基本処理と基本的に

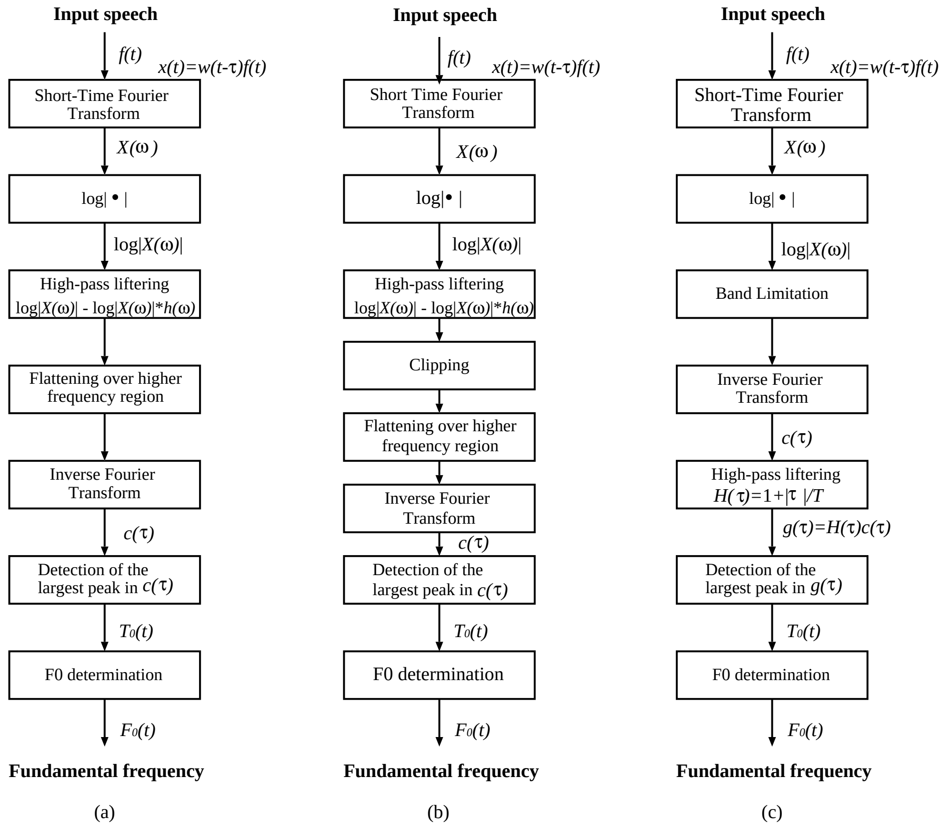

図 8　ケプストラム処理を利用した基本周波数推定方法のブロックダイアグラム．(a) Noll のケプストラム法（基本手法），(b) クリップストラム法，(c) 改良ケプストラム法．

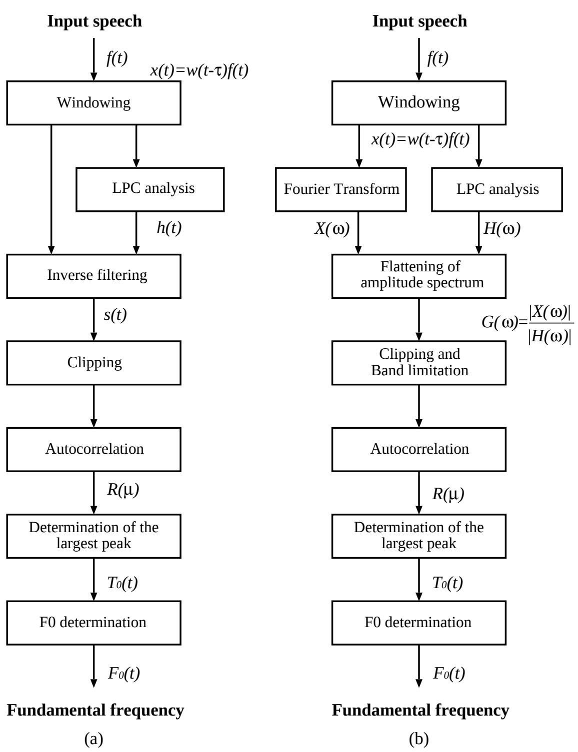

図 9　LPC 法を利用した基本周波数推定のブロックダイアグラム．(a) LPC 残差法，(b) LPC-SIFT 法．

は同じで，対数振幅スペクトルにおける帯域制限を設けたり，フィルタの伝達特性を取り除くために高域通過フィルタに対応するリフターを利用して，音源信号に対応するケプストラム $C_{A,\mathrm{src}}(q,\tau)$ に対して，そのピークとなるケフレンシーを求めることで基本周期，基本周波数を推定している．また，処理法によっては，このピークに荷重関数をかけてピークを強調し，より頑健にピーク位置のケフレンシーを求めるものもある．

3.4　LPC 法を利用した推定方法

図 9 に，LPC 法を利用した基本周波数推定法 [1-4] を示す．これには大別して二種類の方法がある．一つは，図 9(a) に示す，LPC による最小位相特性を有する声道伝達特性 $X_{\mathrm{filt}}(\omega,\tau)$ (or $h(t)$) の推定による音源情報 $s(t)$ の周期性の推定方法である．これは，直接には，LPC 残差信号に対する自己相関法でもある．もう一つは，声道伝達特性の逆フィルタリングを利用した，音源特性 $X_{\mathrm{src}}(\omega,\tau)$ (or $G(\omega)$) の推定およびその周期性を検出するものである．いずれも，音源／フィルタ特性の混在した情報から声道特性を逆推定し，それを利用して音源情報のみを抽出し，従来ある推定法をそれに適用することで基本周波数推定を狙ったものである．

3.5　調波性を利用した推定方法

図 10 に，部分調波性の荷重和（SHS：Sub-Harmonic Summetion）を利用した基本周波数推定法 [29] を示す．この方法は，対数周波数 $\log\omega$ における対数振幅スペクトル $\log|X(\log\omega,\tau)|$ の（部分的）調波性の荷重和を検出することで，基本周波数を推定する方法である．基本的に，Comb フィルタ法と類似する手法であるが，その調波関係を全て見るのではなく，基本周波数のところからある範囲までを見て，その荷重和から推定す

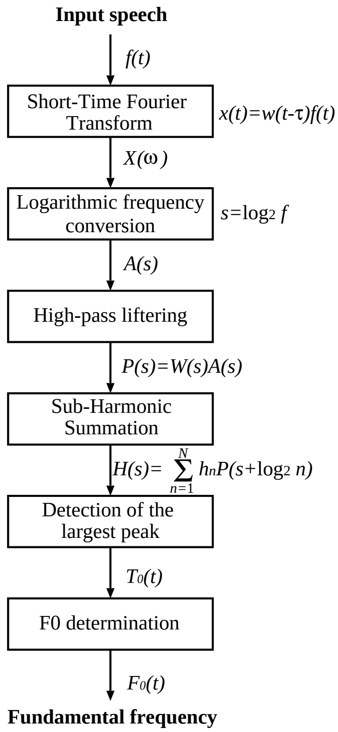

図 10　部分調波性を利用した基本周波数推定法のブロックダイアグラム．

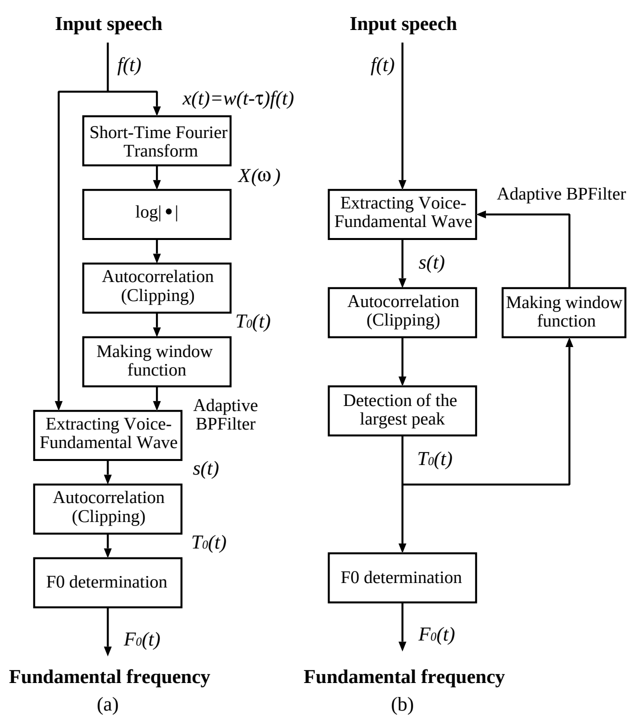

図 11　基本波フィルタリング法を利用した基本周波数推定法のブロックダイアグラム．

るものである．荷重の取り方にも数種類の決め方があるが，おそらく声道フィルタの形状に類似するものが適切であると思われる．

3.6　基本波フィルタリング法を利用した推定方法

図 11 に，基本波フィルタリング法 [27, 28] を利用した基本周波数推定法を示す．この方法は，初段で推定された基本周波数を元にそれを中心とする適応的な帯域通過フィルタを利用して基本波付近の成分を抽出し，それに対し，再度，自己相関法等を利用して基本周波数を推定をするものである．これには二種類の構成があり，一つは図 11(a) に示すようなフィードフォワード形を取るもので，もう一つは図 11(b) に示すようなフィードバック形を取るものである．いずれも初段の推定結果に最終結果が依存する形となる．

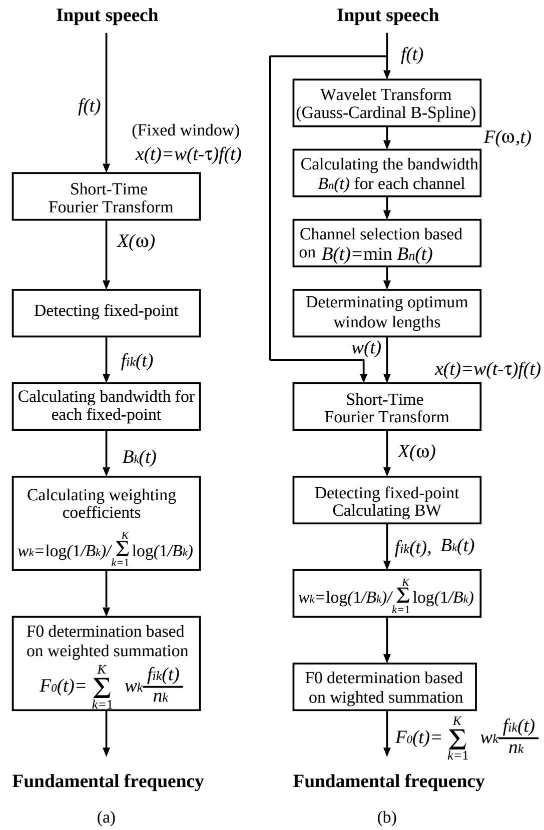

図 12　IFHC 法のブロックダイアグラム．(a) 事前に窓幅を固定して推定する方法，(b) 帯域幅方程式により適切に窓幅を決めて推定する方法．

3.7　瞬時周波数を利用した推定方法

2.4 節で説明した瞬時周波数を利用した基本周波数推定法 [32-39] が報告されている．中でも代表的なものとして，TEMPO 法が知られている．これは，瞬時周波数 $f_i(\omega, \tau)$ を利用することで，調波周波数（基本波の倍音周波数）におけるフィルタ出力での安定な不動点を抽出することで，基本周波数を容易に推定できるというものである．この TEMPO 法は，更に，2.5 節で説明した群遅延 $\theta(q, \tau)$ からの最小位相成分 $\theta_{\min}(\omega, \tau)$ の遅延を補償することで，各瞬時における基本周波数の遅延を補正している [36].

この他，IFHC（Instantaneous Frequency of Harmonic components）法 [38, 39] も提案されている．これは TEMPO 法をベースとしているが，違いは，(i) 帯域幅方程式を利用した最適窓関数の決定，(ii) 瞬時周波数 $f_i(\omega, \tau)$ における不動点の抽出，(iii) 調波性に基づいた不動点の抽出である．図 12 に IFHC の基本処理のブロックダイアグラムを示す．図 12(a) は，窓関数を事前に選択した場合（基本的には TEMPO 法と同じ）の処

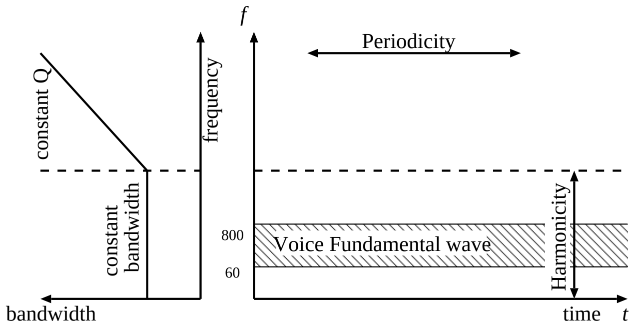

図 13　定 Q ／定帯域フィルタバンクの帯域幅の配置.

理を示し，図 12(b) は，窓関数を帯域幅方程式により決定して (a) の処理に適用するものである．

3.8　瞬時振幅を利用した方法

瞬時周波数の他，瞬時振幅を利用した基本周波数推定法もある．これは，2.2 節で説明した短時間 Fourier 変換や wavelet 変換を利用して，2.2.4 節で説明した瞬時振幅を特徴とし，この中に見られる調波性を検出して基本周波数を推定するものである．例えば，著者らによって提案された手法としては，定 Q フィルタバンクにおける瞬時振幅 $|F(a,b)|$ でのスケール軸に見える調波性（対数的な配置）を検出し，これを自己相関関数や Comb フィルタを利用して基本周波数を推定するものである [40, 41]．

また，高い周波数帯域での瞬時振幅には，時間方向での周期性も見られるため，同様に時間方向に対する自己相関処理より周期性を検出することでも基本周波数を推定することができる．

上記の 2 点を組み合わせた総合的な処理として，PHIA (Periodicity and Harmonicity using Instantaneous Amplitude) 法 [42-44] がある．これは，両者で推定された基本周波数に対し，推定の確からしさを定義し，これを Dempster の結合則を利用して最終結果を推定するものである．この処理では，図 13 に示す 2 種類のフィルタバンクを利用する．二つの推定処理に関しては，調波性に対して定帯域幅フィルタバンクを，周期性に対して定 Q フィルタバンクを利用して，基本周波数候補を推定している．

4.　基本周波数推定法の評価

表 2 に，前節で紹介したものも含め、代表的な基本周波数推定法のアルゴリズムの特徴を示す．ここでは，その中でも精度の高い，次の 16 個の推定法について評価シミュレーションを行うことで，耐残響特性を調べる．

1．時間波形処理
　（1）　自己相関法
　（2）　複数窓長を利用した自己相関法
　（3）　AMDF 法（短時間窓処理）
2.　短時間 Fourier 変換を利用した処理
　（1）　自己相関法（対数振幅スペクトル）
　（2）　Comb フィルタリング法
　（3）　Lag 窓法
　（4）　リフター法
3.　ケプストラム法
　（1）　Noll のケプストラム法（基本処理）
　（2）　改良ケプストラム法
4.　LPC 法
　（1）　残差信号に対する処理
　（2）　SIFT 法
5.　部分調波性の荷重和（SHS）を利用した方法
6.　基本波フィルタリング
7.　TEMPO 法
8.　IFHC 法
9.　PHIA 法

4.1　評価用データ

評価シミュレーションに用いた音声データは，音声と EGG が同時収録された音声データベース [38, 39] を用いた．このデータベースでは，男女各 14 名の発話による 30 文章中（合計 840 文）で構成されるが，本稿ではそのうちの 1 文章（「非常口はどこですか」）を利用した．なお，音声データおよび EGG 波形は，サンプリング周波数 48 kHz で収録されているが，本稿では 20 kHz にダウンサンプリングして利用した．

次に本稿で取り扱う残響特性の加工方法については，次の二つの人工的な残響インパルスを利用した．一つは通常の第 K 次反射波エコーで構成されるインパルス応答 $h_M(t)$，もう一つは指数減衰形の非最小位相インパルス応答 $h_N(t)$ である．

$$h_M(t) \;=\; d(t) + \sum_{k=1}^{K} a_k d(t - k\cdot T_R/100) \tag{53}$$

$$h_N(t) \;=\; a\exp\left(6.9t/T_R\right)\cdot n(t) \tag{54}$$

ただし，次数 $K = 50$, $d(t)$ は Direc のデルタ関数

$$d(t) = \begin{cases} 1, & t = 0 \\ 0, & \text{otherwise} \end{cases} \tag{55}$$

であり，$n(t)$ はガウス性ランダム雑音，T_R はパワーが 60 dB 減衰するときの残響時間である．パラメータ a は，パワーを調整するものであるため，ここでは，

$$a = \sqrt{1/\int_0^T \exp(-13.8t/T_R)dt} \tag{56}$$

とした．また，評価条件として，$h_M(t)$ の残響時間 $T_R = 0.5$, 1, 3, 5, 10, 20 ms の合計 6 種類とした．また、$h_N(t)$ の残響時間を $T_R = 0.05$, 0.1, 0.3, 0.5, 1.0, 2.0 s の合計 6 種類とした．

4.2　評 価 尺 度

各推定方法の推定精度を評価するためのリファレンス（$F_0(t)$）として，前節で述べた音声データベース中の EGG 信号に対し，現在もっとも信頼が高く精度が高い推定方法である TEMPO 法 [34] で推定されたものを利用する．また，各推定法で推定した基本周波数を $\hat{F}_0(t)$ と置き，次の二つの評価尺度（正当率と SNR）を利用して，各推定法の推定精度を評価する．

$$\text{Correctness} \;=\; (N_C/N_V)\times 100 \tag{57}$$

$$\text{SNR} \;=\; 20\log_{10}\frac{\int F_0(t)dt}{\int(F_0(t)-\hat{F}_0(t))dt} \quad \text{(dB)} \tag{58}$$

表 2　基本周波数推定法（アルゴリズム）で利用された特徴.

アルゴリズム	処理領域	周期性	調波性	フィルタ形状	位相特性	放射特性	特徴	引用文献
時間波形処理								
(1) ゼロ交差	時間	o	x	x	x	x	$x(t,\tau)$	[3, 5, 6]
(2) ピーク検出	時間	o	x	x	x	x	$x(t,\tau)$	[3, 7]
(3) 自己相関	時間	o	x	x	x	x	$x(t,\tau)$	[1, 3, 8]
複数窓長を利用した自己相関法	時間	o	x	x	x	o	$x(t,\tau)$	[9]
AMDF								
(1) 短時間窓処理	時間	o	x	x	x	x	$x(t,\tau)$	[3, 10]
(2) LPC 残差信号	時間	o	x	o	x	x	$s(t,\tau)$	
短時間 Fourier 変換								
(1) 自己相関	Frequency	x	o	x	x	x	$\|X(\omega,\tau)\|$	[1, 2, 3, 4, 19, 20]
(2) Comb フィルタリング	Freq.	x	o	x	x	x	$\|S(\omega,\tau)\|$	[22, 23, 24, 25]
(3) Lag 窓	Freq./Quef.	x	o	x	x	x	$\|S(\omega,\tau)\|$	[21]
(4) リフター	Frequency.	x	o	x	x	x	$S(\omega,\tau)\|$	[1, 2, 3]
ケプストラム法								[3, 11]
(1) Noll のケプストラム法	ケフレンシー	o	x	x	x	x	$C_A(q,\tau)$	[1, 2, 3, 4, 12, 13]
(2) クリップストラム	ケフレンシー	o	x	x	x	x	$C_A(q,\tau)$	[1, 2, 3, 4, 14]
(3) 改良ケプストラム	ケフレンシー	o	x	x	x	x	$C_A(q,\tau)$	[1, 2, 3, 4, 15, 16, 17]
LPC 法								
(1) 残差信号	時間	o	x	x	x	x	$s(t,\tau)$	[1, 2, 3, 4]
(2) SIFT 法	周波数	x	o	x	x	x	$\|S(\omega,\tau)\|$	[18]
部分調波の荷重和 (SHS)	周波数	x	o	x	x	x	$\log\|X(\omega,\tau)\|$	[29]
基本波フィルタリング	時間	o	o	x	x	x	$s(t,\tau)$	[27, 28]
TEMPO (Previous)	周波数	x	x	x	o	x	$f_i(\omega,\tau)$	[33]
TEMPO	周波数	x	x	x	o	x	$f_i(\omega,\tau)$	[34, 35, 36]
IFHC 法	周波数	x	o	x	o	x	$f_i(\omega,\tau)$	[33]
定 Q フィルタバンク								
(1) Comb フィルタリング	周波数	x	o	x	x	x	$\|X(\omega,\tau)\|$	[40, 41]
(2) 自己相関法	Time	o	x	x	x	x	$\|X(\omega,\tau)\|$	
定帯域フィルタバンク								
(1) Comb フィルタリング	周波数	x	o	x	x	x	$\|X(\omega,\tau)\|$	
(2) ACorr	周波数	x	o	x	x	x	$\log\|X(\omega,\tau)\|$	
(3) SHS 法	周波数	x	o	x	x	x	$\log\|X(\omega,\tau)\|$	
PHIA 法	時間/周波数	o	o	x	x	x	$\|X(\omega,\tau)\|$	[42, 43, 44, 45]
提案法	時間/周波数	o	o	o	o	—	$s(t,\tau)$	—

ただし，N_V は有声区間と推定された区間長，N_C は推定誤差 $F_0(t)-\hat{F}_0(t)$ が 5 % ないし 10%のときの区間長での割合を示す．ここで，正答率は，有声区間内で，ある指定された誤差率の範囲内でどれだけ正しく基本周波数を推定できたかを示し，SNR はそのときの外形の誤差を示す．SNR は高ければ高いほど高い精度であることを指す．

4.3 評 価 結 果

図 14 に波形信号の自己相関法による推定結果を示す．図 14(a), (b) にそれぞれ，誤差率 5%と 10%内の基本周波数の推定に対する正答率を示す．横軸は残響時間 T_R (s) を示すが，最小位相特性をもつ残響インパルス応答 $h_M(t)$ に対する残響時間は表示時間の 1/100 である．また，図 14(c)，(d) に残響時間に対する平均 SNR と標準偏差を示す．これらの結果から，残響がないクリーンなときは比較的よい精度で基本周波数を推定できており，また最小位相特性をもつ残響インパルス応答 $h_M(t)$（○の実線）に対しては，比較的頑健に基本周波数を推定できていることがわかる．しかし，非最小位相特性をもつ残響インパルス応答 $h_N(t)$（△の破線）に対しては，$T_R=0.1$ s から急激に推定精度が低下しており，$T_R=2.0$ s のところでは既に正答率が元の精度の半分以下に達していることがわかる．

図 15 に波形信号の複数窓長を利用した自己相関法による推定結果を，図 16 に波形信号に対する AMDF 法による推定結果を示す．図のフォーマットは，図 14 に示したものと同じである．いずれの結果をみても，最小位相特性を有する残響インパルス応答 $h_M(t)$ に対しては，比較的頑健に基本周波数を推定できるのに対し，非最小位相特性をもつ残響インパルス応答 $h_N(t)$ に対しては，ほとんど推定できておらず，そのときの精度は $h_M(t)$ のときの結果の半分以下になっていることがわかる．

以上の結果から，時間波形における基本周波数推定法は，いずれも残響の影響を直接受けた音声から推定しなければならないため，推定精度も直接，残響の影響を受けており，残響に関して頑健な推定法であるとはいい難い．

次に，図 17 に短時間 Fourier 変換における対数振幅の自己相関法による推定結果を示す．図のフォーマットはこれまでのものと同じである．また，横軸は残響時間 T_R (s) を示すが，最小位相特性をもつ残響インパルス応答 $h_M(t)$ の残響時間は，表示されている T_R の 1/100 である．これらの結果から，残響がないクリーンなときは比較的よい精度で基本周波数を推定できており，また最小位相特性をもつ残響インパルス応答 $h_M(t)$（○の実線）に対しては，比較的頑健に基本周波数を推定できてい

ることがわかる．しかし，非最小位相特性をもつ残響インパルス応答 $h_N(t)$（△の破線）に対しては，$T_R = 0.1$ s から急激に推定精度が低下しており，$T_R = 2.0$ s のところでは既に正答率が元の精度の半分以下に達していることがわかる．

図 18 に短時間 Fourier 変換における Comb フィルタ法による推定結果を，図 19 に短時間 Fourier 変換における Lag 窓法による推定結果を，図 20 に短時間 Fourier 変換におけるリフター法による推定結果を示す．いずれの結果も，図 17 と同様のものとなっており，非最小位相特性をもつ残響インパルス応答 $h_N(t)$ に対しては，残響時間が長いときほとんど基本周波数を正しく推定できていないことがわかる．

図 21 に Noll のケプストラム法による推定結果を，図 22 に改良ケプストラム法による推定結果を示す．図のフォーマットはこれまでと同じものである．これらの結果から，残響がないクリーンなときは比較的よい精度で基本周波数を推定できており，また最小位相特性をもつ残響インパルス応答 $h_M(t)$（○の実線）に対しては，比較的頑健に基本周波数を推定でき，非最小位相特性をもつ残響インパルス応答 $h_N(t)$（△の破線）に対しては，これまでの結果と同様に推定精度が著しく低下していることがわかる．しかし，推定誤差率 10%内における正答率に着目すると，これまでの方法による結果と比較すると 10%近い正答率の向上が見られる．これは，ケプストラム処理が Homomorphic 変換であるため，同じ成分をもつ等化な遅延（例えばエコーのようなもの）については，元のものと遅延成分を同じ成分として取り扱えることができるためである．この手法による結果が決して最良なものであるとは思わないが，残響に頑健な基本周波数推定法を確立するための一つの手がかりとして，有効な手法であると考えられる．

次に，図 23 に LPC 法（残差信号）による推定結果を，図 24 に LPCF 法（SIFT）による推定結果を示す．この結果もこれまでの結果と同様の傾向を示しており，非最小位相特性をもつ残響インパルス応答 $h_N(t)$ のとき著しく推定精度が低下していることがわかる．同じ LPC 法の中では，LPC 残差を直接利用するものがまだ推定精度が他のものよりは良くなっていることがわかる．これは，ケプストラム法を利用したものと合わせて考えると，音源・フィルタモデルを考えたときの音源情報のみを抽出し，それに対して基本周波数を推定したほうがよいことを示唆している．

次に，図 25 に SHSF 法による推定結果を，図 26 に基本波法による推定結果を示す．いずれも調波性，あるいは調波内のうちの基本波に着目して推定する方法であるが，いずれもこれまでの結果と同様，非最小位相特性をもつ残響インパルス応答 $h_N(t)$ のとき著しく推定精度が低下していることがわかる．特に，基本波フィルタリング法では，残響時間が長いとき，これまでの中で一番精度が低下していることがわかる．これはおそらく初段の基本周期の推定に失敗し，それの影響を受けた状態で再度推定していることに起因していると思われる．

図 27 に TEMPO 法による推定結果を，図 28 に IFHC 法による推定結果を示す．いずれも瞬時周波数を利用するものであり，後者はこれに調波性を組み込んだものである．TEMPO 法では残響がないとき，非常によく基本周波数を推定できていることがわかる．しかし，いずれの手法も残響時間が長くなるにつれ，推定精度が著しく低下しており，TEMPO 法では，基本波フィルタリングと同程度に精度が低下している．

同じく，図 29 に PIHA 法よる推定結果を示す．この手法は雑音に非常に頑健な手法の一つであり，著者らによって提案された瞬時振幅・瞬時周波数における周期性・調波性を利用した基本周波数推定法の初段の推定法である．この手法は，PHIA 法を初段の推定法として利用し，これを元に雑音抑圧した後，TEMPO 法を利用して基本周波数を推定するものであるが，図 29 の結果を初段の推定結果として，最終的に基本周波数を推定するため，顕著な改善を期待できないかもしれない．

全体的に，図 14〜29 を通して結果を眺めると，残響特性に頑健と考えられる特徴は，(1) Homomorphic 変換により遅延情報を取り除かれた信号における周期性あるいは調波性と，(2) 音源フィルタモデルを仮定して抽出された音源情報における周期性あるいは調波性である．そこで，本稿では，残響に頑健な基本周波数推定法の方略として，上記 2 点を同時に利用した基本周波数推定法を考える．

5. 残響に頑健な基本周波数推定法の提案

5.1 提案法のアルゴリズム

音源情報の抽出と，信号成分の遅延情報の分離を同時に行うことが可能な分析法として，2.3 節で説明した複素ケプストラム分析が有効であると考えられる．理由として次のことがあげられる．もともとケプストラム分析では，複素スペクトルの積の表現を和の表現に変換できるものであるため，音源情報と声道フィルタ特性がケフレンシー領域で重複していないと仮定すれば，容易にこれらを切り分けて扱うことができるものである．また，ケプストラム分析は Homomorphic 分析であり，信号成分の一定な遅延を同じ信号成分として取り扱うことができるため，最小位相特性をもつようなインパルス応答による影響を取り除くことできる．

そのため，図 22 等の結果では，比較的良好な推定精度を示していた．しかし，これらのケプストラムを利用した手法では，振幅ケプストラムのみを利用しており，位相情報を取り扱っていなかったため，時間領域における信号の周期性の手がかりを直接利用していなかった．

本稿では，音源フィルタモデルを仮定し，複素ケプストラムを利用して，音源情報をもつ複素ケプストラムと声道フィルタ特性をもつ複素ケプストラムを取り扱い，また同時にその音源信号の周期性・調波性も取り扱う．本来は，残響の影響を考慮して基本周波数を推定することから，

$$s(t) * h(t) * h_M(t) \qquad \text{or} \qquad s(t) * h(t) * h_N(t) \qquad (59)$$

という二つの伝達特性（一つは声道伝達特性ともう一つは残響の伝達特性）を取り除いて音源情報のみから基本周波数を推定することになる．しかし，$h_M(t)$ ないし $h_N(t)$ を逆推定してから基本周波数を推定することがかなり困難であるため，$h(t)$ と，$h_M(t)$ ないし $h_N(t)$ を切り分けずにそれらの影響を同時に

考慮することにする．声道フィルタ $h(t)$ の伝達特性は，ケフレンシー領域では，低ケフレンシー領域に集中すると考えられるため，ここでは，図 3 に示すように，リフター処理にてこららを切り分けることにする．

残響の伝達特性は，単純に考えると信号の周波数成分に対して遅延をかけるものと考えられる．そのため，Homomorphic 分析の特徴を生かせば，音源信号の特性そのものの遅延を容易に取り除けることになる（特にエコーのような場合にこの分析の能力を発揮できる）．これ以外の遅延を担う情報が，非最小位相成分，つまり全域通過位相成分にあると考えれば，最小位相成分だけを抽出することでこの影響を取り除くことができる．2.3.2 節で述べたように，複素ケプストラムから容易に，最小位相ケプストラムを算出できることから，結果として複素ケプストラム分析を利用することが，最有力な分析法であるといえる．

この考えに基づいた基本周波数推定法のアルゴリズムを図 31 に示す．おおよそこれには 2 種類の方法が考えられる．いずれも複素ケプストラム分析を利用して，最小位相成分を有する信号を観測信号から抽出し（残響による位相の影響を取り除き），最小位相の複素ケプストラムに関してケフレンシーにおいて，リフタリング処理により声道フィルタ特性を取り除くことで，音源情報のみを抽出する．これをベースとなる初段の処理とすると，後段の処理として，音源信号の周期性を取り扱うのか，調波性を取り扱うのかによって様々な手法が考えられる．音源の周期性を自己相関法によって取り出すものが図 31(a) のアルゴリズムであり，音源の調波性をスペクトル領域にて自己相関法で取り扱うものが図 31(b)，Comb フィルタリング処理で取り扱うものが図 31(c)，SHS 法で取り扱うものが図 31(d) のアルゴリズムである．

5.2 推定法の評価

図 32 に，音源フィルタモデルを仮定し，複素ケプストラムを利用して得られた音源信号の周期性を自己相関法にて抽出した場合の基本周波数推定の結果を示す．図のフォーマットはこれまでのものと全く同じである．横軸は，残響時間を示すが，最小位相特性をもつ残響インパルス応答 $h_M(t)$ の場合，残響時間は表示された値の 1/100 になっている．結果を見ると，残響の影響がないときの基本周波数の精度は，TEMPO 等に比べて精度としては若干劣るものの，両残響特性に対して，残響時間が長くなっても精度が半分以下に低下することなく，おおよそフラットでかつ正答率 60%以上を保持していることがわかる．この結果は，絶対的に良い推定精度とはいえないかもしれないが，本稿で紹介した全手法の結果と比較すると，明らかに高い推定精度を得ていることがわかる．

図 33 に，複素ケプストラム分析を用いた音源フィルタモデルにおいて，音源スペクトルの自己相関法にて推定された結果を，図 34 に音源スペクトルの調波性を Comb フィルタリングにて抽出した場合の推定結果を，図 35 に音源スペクトルの SHS 法にて推定された結果を示す．これらの結果では，図 32 と同様に，残響時間に対してあまり影響を受けずに基本周波数を推定できることがわかる．特に，図 34 に示す Comb フィルタリングの場合，音源信号の調波性をより良く抽出できているため，図 32 の結果と同程度に，残響に頑健に基本周波数を推定できていることがわかる．

これらの結果を踏まえると，残響にロバストな基本周波数推定法の方略として，音源フィルタモデルを仮定したときの音源情報のみを扱い，同時に残響の影響による遅延を取り除くことを考慮した上で，音源情報の周期性あるいは調波性を利用して基本周波数を推定することが最も有効であることがわかった．残響の影響が少ないときの絶対的な推定精度がまだ十分高いとはいえないが，本手法をベースとして，TEMPO 法などで有効である瞬時周波数を更に利用するとこで，残響環境下において，残響にロバストで高精度な基本周波数推定法を確立できるものと考えられる．

6. まとめ

本稿では，残響にロバストな基本周波数推定法を確立するための基礎検討として，数学的な記述の準備を踏まえた上で，これまでに知られている代表的な基本周波数推定法を紹介した．その後で，最小位相特性を有する残響インパルス応答（エコー）と，非最小位相特性を有する残響インパルス応答の，二種類の残響特性を利用して，これらの手法の残響に対する頑健性を調査した．その結果，ほとんどの結果は，エコーのような最小位相特性をもつ残響特性について，それほど大きく影響を受けることなく，ある程度の精度をもって基本周波数を推定できることがわかった．しかし，非最小位相特性をもつ残響特性に関しては，いずれの手法も残響に大きく影響を受け，ほとんど正しく推定できないことがわかった．特に，ほとんどの手法で，最小位相特性をもつ残響特性に対する基本周波数推定精度に対し，非最小位相特性もつ残響の場合，約半分以下までその推定精度を落としてしまうことがわかった．

各手法で利用されている特徴表現を比較したところ，(1) Homomorphic 変換により遅延情報を取り除かれた信号における周期性あるいは調波性と，(2) 音源フィルタモデルを仮定して抽出された音源情報における周期性あるいは調波性が，残響特性に関して頑健で有効な特徴であることがわかった．そこで，残響に頑健な基本周波数推定法の方略として，音源フィルタモデルを仮定し，これらの特徴を複素ケプストラム分析を利用して，音源信号の周期性あるいは調波性を検出して基本周波数を推定する方法を提案した．同様の評価実験を行ったところ，本稿で調査したすべての手法（16 個）と比べて，本手法が残響にロバストで，かなり高い推定精度をもつことがわかった．この手法には，まだいくつか解決しなければならない課題（例えば，クリーンなときの推定精度の向上など）があるが，残響にロバストな推定法のベースモデルとして，非常に有効なものであると考えられる．

謝　辞

本研究は，科学研究費補助金（No.14780267) の援助を受けて行われた．

文　献

[1] Hess, W. J. “Pitch Determination of Speech Signals,”

Springer-Verlag, New York 1983.

[2] Hess, W. J. "Pitch and Voicing Determination," in Adcanced in speech signal processing, ed. S. Furui and M.M. Sondhi, pp. 3-48, Marcel Dekker. Inc. New York 1992.

[3] Furui, S., ディジタル音声処理，東海大学出版会，1985 (in Japanese).

[4] Suzuki, H., "A story of old-and news of pitch extracton in speech technology," J. Acoust. Soc. Jpn. vol. 56, no. 2 pp. 121-128, 2000 (in Japanese)

[5] Gold, B. and Rabiner, L., "Parallel processing techniques for estiming pitch periods of speech in the time domain," J. Acoust. Soc. Am., vol. 46, no. 2, pp. 442-448, Aug. 1969.

[6] Schroeder, M. R., "Period histogram and product spectrum: new methods for fundamental frequency measurement," J. Acoust. Soc. Am., vol. 43, no. 4, pp. 829-834, April 1968.

[7] Howard, D. M., "Peak-picking fundamental period estimation for hearing prostheses," J. Acoust. Soc. Am., vol. 86, no. 3, pp. 902-910, Sept. 1989.

[8] de Cheveign'e, A., Kawahara, H., "Yin, a fundamental frequency estimator for speech and music," J. Acoust. Soc. Am., vol. 111, no. 4, pp. 1917-1930, 2002.

[9] Takagi,T., Seiyama, N., and Miyasaka, E., "A Method for pitch extraction of speech signal using autocorrelation functions through multiple window-length," IEICE vol. J80-A, no. 9, pp. 1341-1350, Sept. 1997 (in Japanese).

[10] Ross, M. J., Shaffer, H. L., Cohen, A., Freudberg, R., and Manley, H. J., "Average magnitude difference function pitch extractro," IEEE Trans. Acoust., Speech, Signal Process. ASSP-22, pp. 353-361, 1974.

[11] Un, C. K., and Yang, S. C. "A pitch extraction algorithm based on LPC inverse filtering and AMDF," IEEE Trans. Acoust., Speech, Signal Process. ASSP-25, pp. 565-572, 1977.

[12] Noll, A. M. "Short-time spectrum and "cepstrum" techniques for vocal-pitch detection," J. Acoust. Soc. Am., vol. 36, no. 2, pp. 226-302, Feb. 1964.

[13] Noll, A. M. "Cepstrum pitch determination," J. Acoust. Soc. Am., vol. 41, no. 2, pp. 293-309, Aug. 1966.

[14] Noll, A. M. "Clipstrum pitch determination," J. Acoust. Soc. Am., vol. 44, no. 6, pp. 1585-1591, July. 1968.

[15] Oppenheim, A. V. "Speech analysis-synthesis system based on homomorphic filtering," J. Acoust. Soc. Am., vol. 45, no. 2, pp. 458-465, 1969.

[16] Kato, S. and Miwa, J., "Pitch detection usign moving average and band-limitation in cepstrum method and its application," Tech. Rep. of IEICE, SP94-95, Feb. 1995.

[17] Oppenheim, A. V. and Schafer, R. W. "Homomorphic analysis of speech," IEEE Trans. Audio, Electroacoust., AU-16, no. 2, pp. 221-226, 1968.

[18] Markel, J., "The SIFT algorithm for fundamental frequency estimation," IEEE Trans. Audio vol. AU-20, pp. 367-377, 1972.

[19] Kobayashi, H. and Shimamura, T., "An extraction method of fundamental frequency using clipping and band limitation on log spectrum," IEICE vol. J82-A, no. 7, pp. 1115-1122, July 1999 (in Japanese).

[20] Kunieda, N., Shimamura, T., and Suzuki, J., "Pitch extraction by using autocorrelation function on the log spectrum," IEICE vol. J80-A, no. 3, pp. 435-443, March 1997 (in Japanese).

[21] Singer, H. and Sagayama, S., "Pitch dependent phone modelling for HMM-based speech recognition," J. Acoust. Soc. Jpn. (E) vol 15, pp. 77-86, 1994.

[22] Nishi, K. and Ando, S., "An optimal comb filter for time-varying harmonics extraction,", IEICE Trans. Fundamentals, vol. E81-A, no. 8, pp. 1622-1627, Aug. 1998.

[23] Nishi, K. and Ando, S., "Uniform-Q comb filter and its time/frequency characteristics – filter architecture for fluctuation error –" IEICE vol. J81-A, no. 2, pp. 152-160, Feb. 2000 (in Japanese).

[24] de Cheveigue, "Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing," J. Acoust. Soc. Am., vol. 93, no. 6, pp. 3271-3290, June 1993.

[25] Miwa, T., Tadokoro, Y., and Saito, T., "The pitch estimation of different musical instruments sounds using comb filters for transcription," IEICE, vol. J81-D-II, no. 9, pp. 1965-1974, Sept. 1998 (in Japanese).

[26] Yanagisawa, K., Tanaka, K., and Yamaura, I., "A detection method of fundamental period using time continuous properties of spectrum envelope," IEICE vol. J83-D-II, no. 11, pp. 2087-2098, Nov. 2000 (in Japanese).

[27] Ohmura, H. and Tanaka, K., "Fine pitch contour extraction by voice fundamental wave filtering method," J. Acoust. Soc. Jpn, vol. 51, no. 7, pp. 509-518, July 1995 (in Japanese).

[28] Sasou, A. and Nakamura, S., "A pitch extraction method using wavelet transform," IEICE A vol. J80-A, no. 11, pp. 1848-1856, Nov. 1997 (in Japanese).

[29] Hermes, D. J., "Measurement of pitch by subharmonic summation," J. Acoust. Soc. Am., vol. 83, no. 1, pp. 257-264, Jan. 1988.

[30] Immerseel, L. M. and Martens, J. P., "Pitch and voiced/unvoiced determination with an auditory model," J. Acoust. Soc. Am., vol. 91, no. 6, pp. 3511-3526, June 1992.

[31] Terhardt, E., Stoll, G., and Seewann, M., "Algorithm for extraction of pitch and pitch salience from complex tonal signals," J. Acoust. Soc. Am., vol. 71, no. 3, pp. 679-688, March 1982.

[32] Brown J. C. and Puckette, M. S., "A high resolution fundamental frequency determination based on phase changes of the Fourier transform," J. Acoust. Soc. Am., vol. 92, no. 2, pp. 662-667, August 1993.

[33] Kawahara, H., Masuda-Katsuse, I., and de Cheveigné, A., "Restructuring speech representations using a pitch-adaptive time-frequency smooting and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," Speech Communication, vol. 27, pp. 187-207, 1999.

[34] Kawahara, H., Katayose, H., Patterson, R. D., and de Cheveigné, A., "Fixed point anaysis of frequency to instantaneous frequency mapping for accurate estimation of f0

and periodicity," Proc. Eurospeech'99, vol. 6, pp. 2781-2784, Sept. 1999.

[35] Kawahara, H. and Zolfaghari, P., "Excitation source information extraction based on fixed-point analysis," Tech. Rep. of IEICE, 1999.

[36] Kawahara, H. and Atake, Y., "Vocal fold closure and speech event detection using group delay, " Tech. Rep. of IEICE, SP99-171, March. 2000.

[37] Abe, T. Kobayashi, T., and Imai, S., "Pitch estimation based on instantaneous frequency in noisy environments," IEICE vol. J79-D-II, no. 11, pp. 1771-1781, Nov. 1996 (in Japanese).

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

[39] Atake, Y., Irino, T., Kawahara, H., Lu, J., Nakamura, S., and Shikano,K., "Robust estimation of fundamental frequency using instantaneous frequencies of harmonic components," IEICE vol. J83-D-II, no. 11, pp. 2077-2086, Nov. 2000.

[40] Unoki, M. and Akagi, M., "A method of extracting the harmonic tone from noisy signal based on auditory scene analysis," IEICE vol. J82-A, no. 10, pp. 1497-1507, Oct. 1999 (in Japanese).

[41] Unoki, M. and Akagi, M., "Signal extraction from noisy signal based on auditory scene analysis," Proc. ICSLP98, vol. 4, pp. 1515-1518, Dec. 1998.

[42] Ishimoto, Y., Unoki, M., and Akagi, M., "A fundamental frequency estimation method for noisy speech based on periodicity and harmonicity," Trans. Tech. Com. Psycho. Physio. Acoust. , H-2000-81, Sept 2000.

[43] Ishimoto, Y., Unoki, M., and Akagi, M., "A fundamental frequency estimation method for noisy speech based on periodicity and harmonicity," Proc. ICASSP2001, 2001.

[44] Ishimoto, Y., Unoki, M., and Akagi, M., "A Fundamental Frequency Estimation Method for Noisy Speech Based on Instantaneous Amplitude and Frequency," Proc. EuroSpeech2001, pp. 2439-2442, Sept. 2001.

[45] Ishimoto, Y., Unoki, M., and Akagi, M., "Fundamental frequency estimation for noisy speech based on instantaneous amplitude and frequency," JAIST Research Report, IS-RR-2005-006, March 2005.

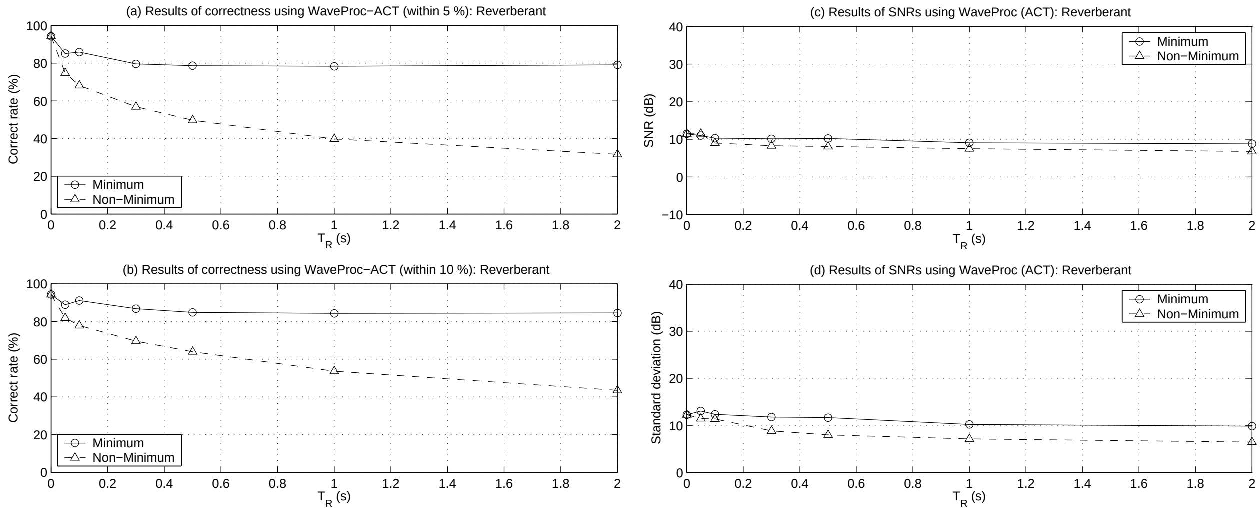

図 14　自己相関法による推定結果．(a) 5%誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差．

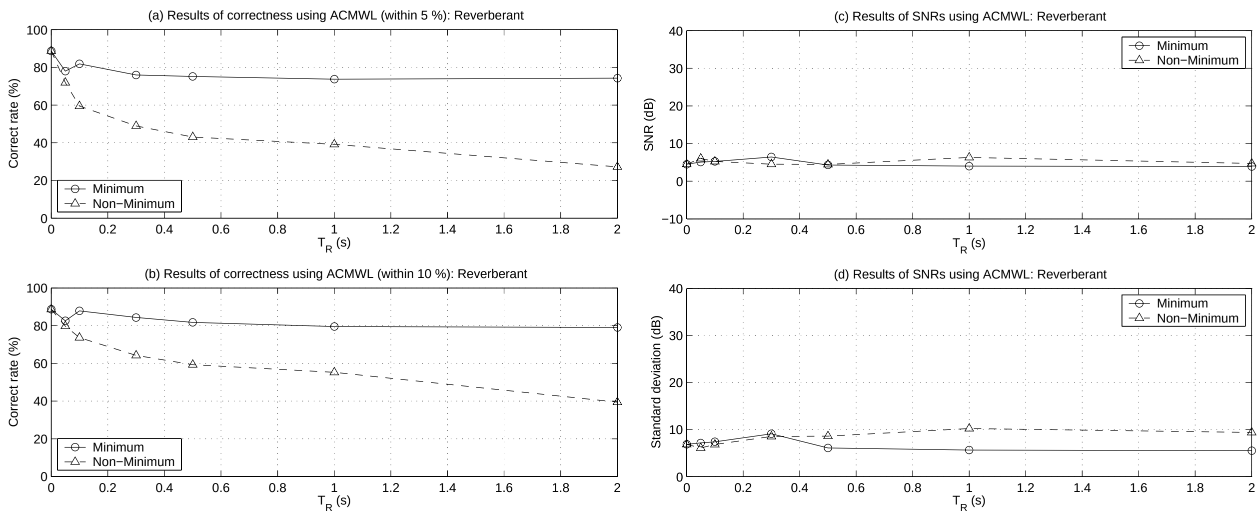

図 15　複数窓長を利用した自己相関（ACMWL）法による推定結果．(a) 5% 誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差．

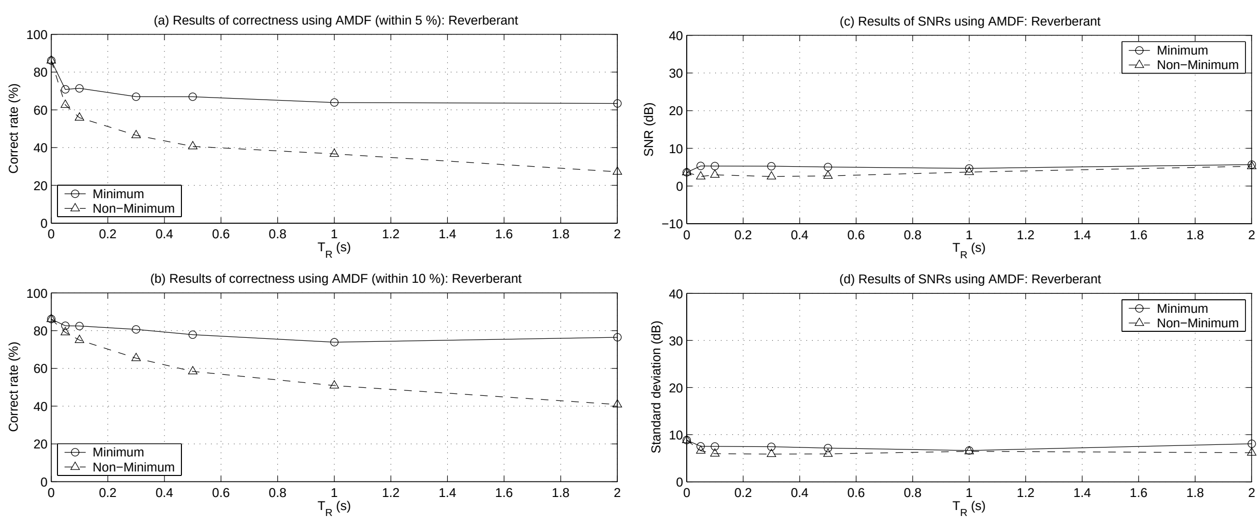

図 16　AMDF 法による推定結果．(a) 5%誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差．

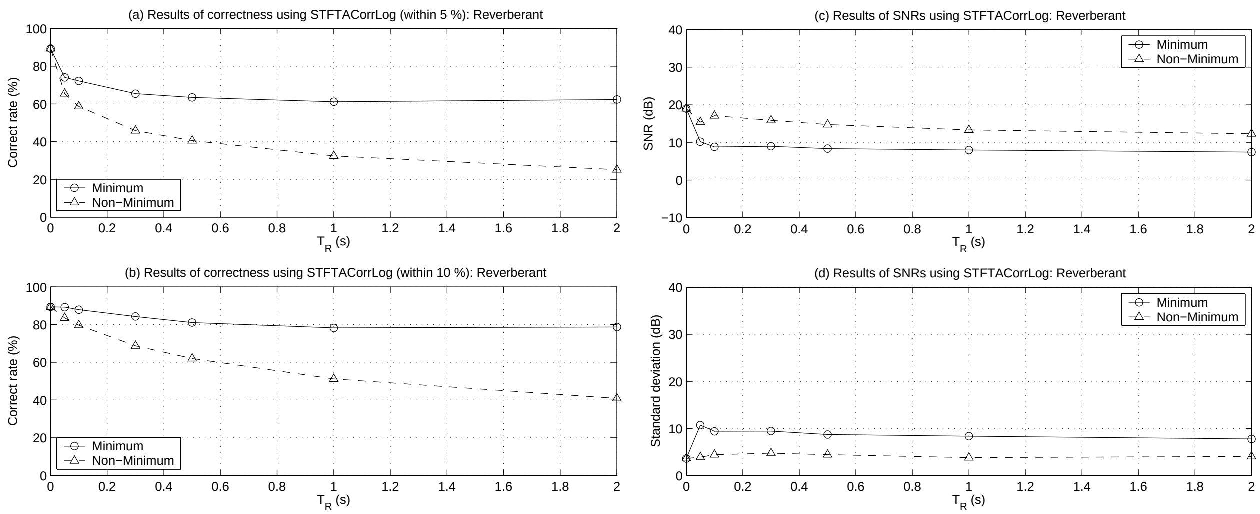

図 17　STFT 法 (Log) による推定結果．(a) 5%誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差．

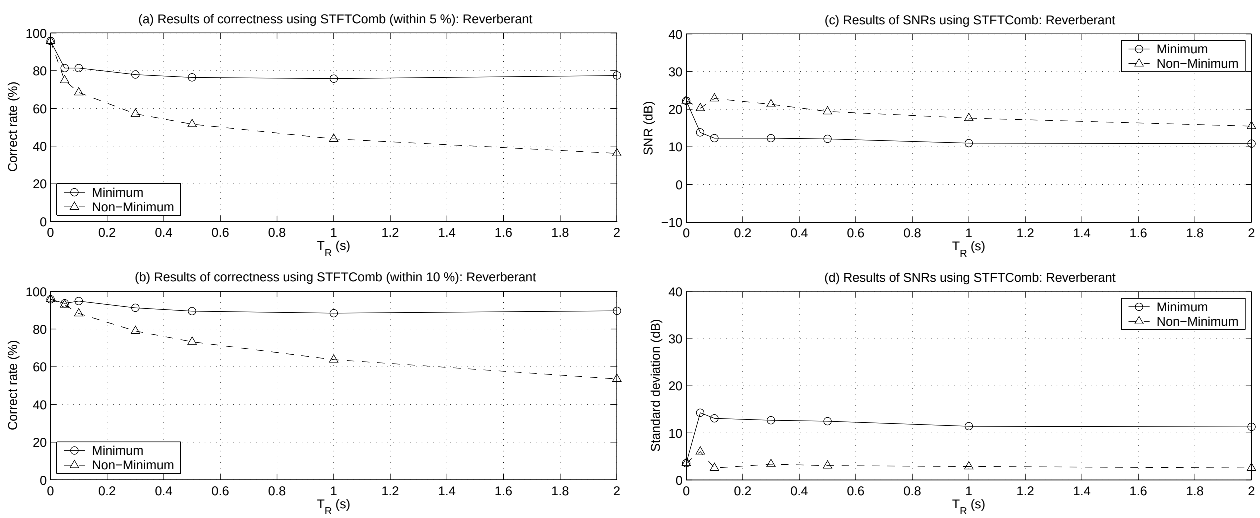

図 18　STFT 法 (Comb フィルタリング) による推定結果．(a) 5%誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差．

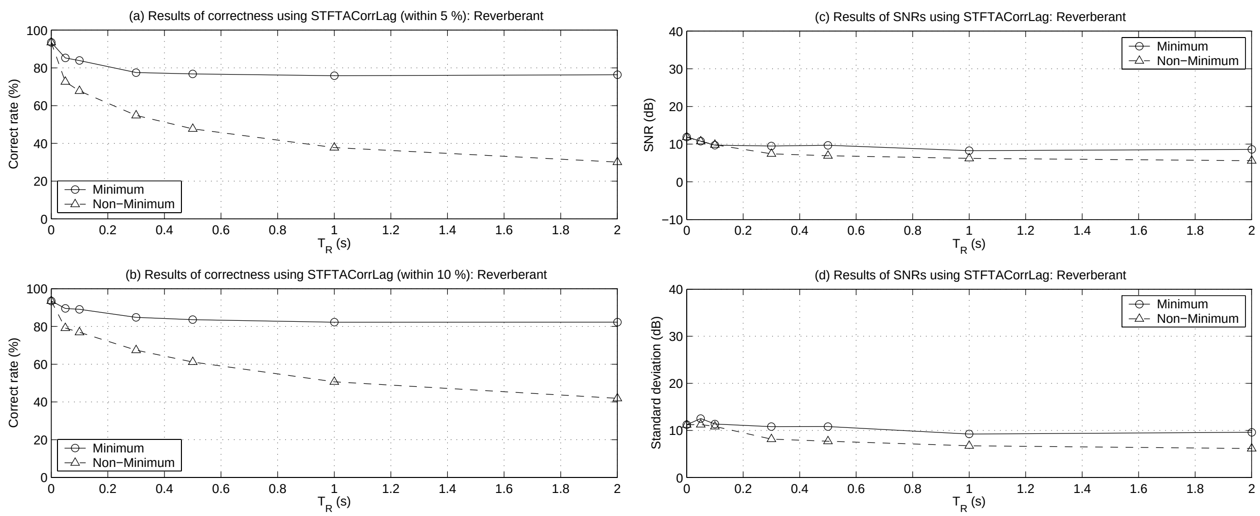

図 19　STFT 法 (Lag) による推定結果．(a) 5%誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差．

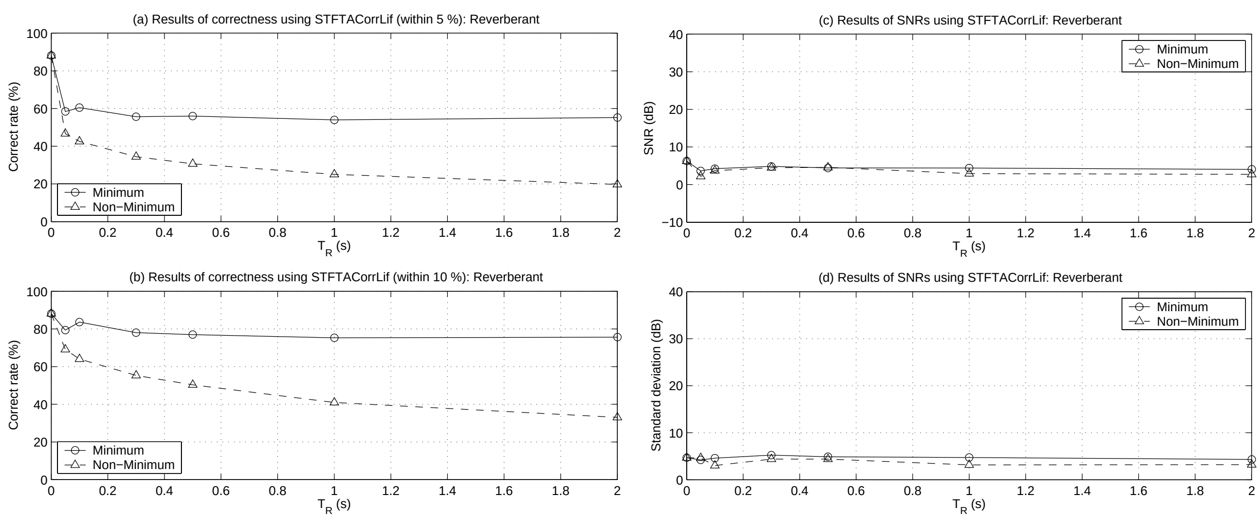

図 20　STFT 法 (Lifter) による推定結果．(a) 5%誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差．

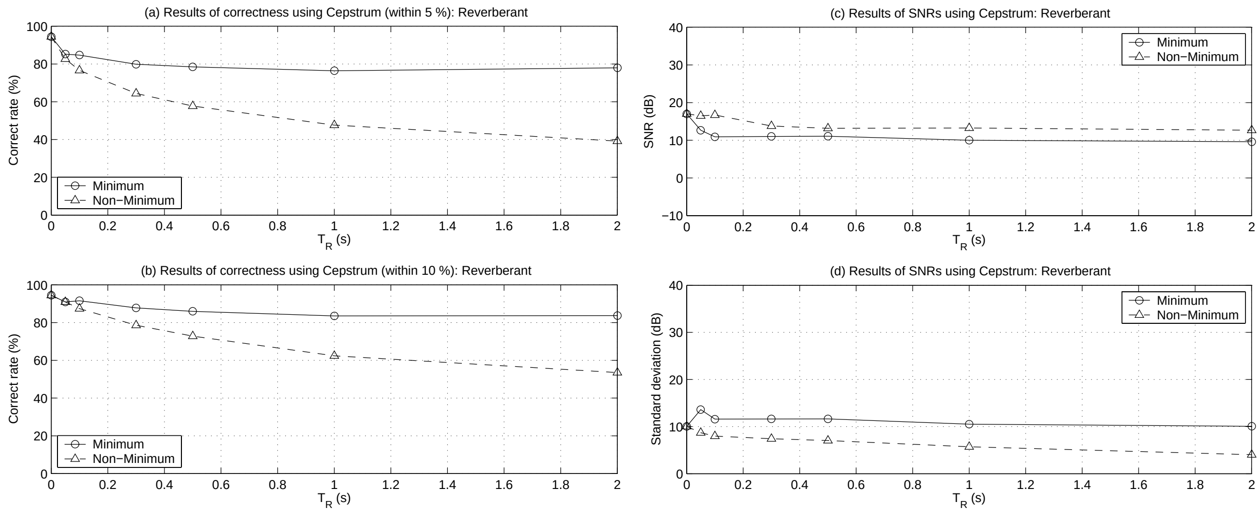

図 21　Cepstrum 法による推定結果．(a) 5%誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差．

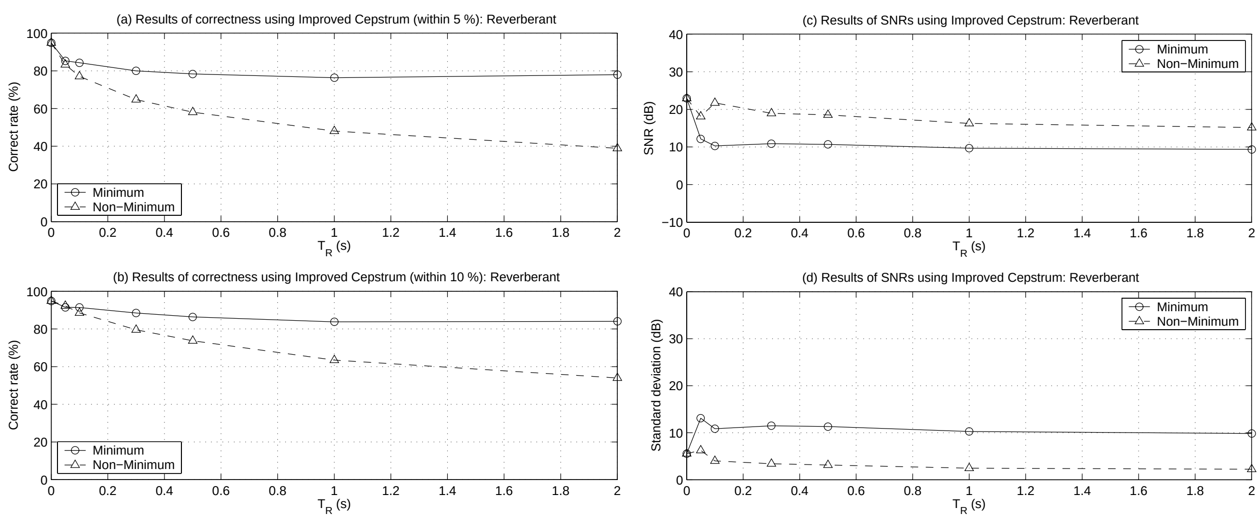

図 22　改良 Cepstrum 法による推定結果．(a) 5%誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差．

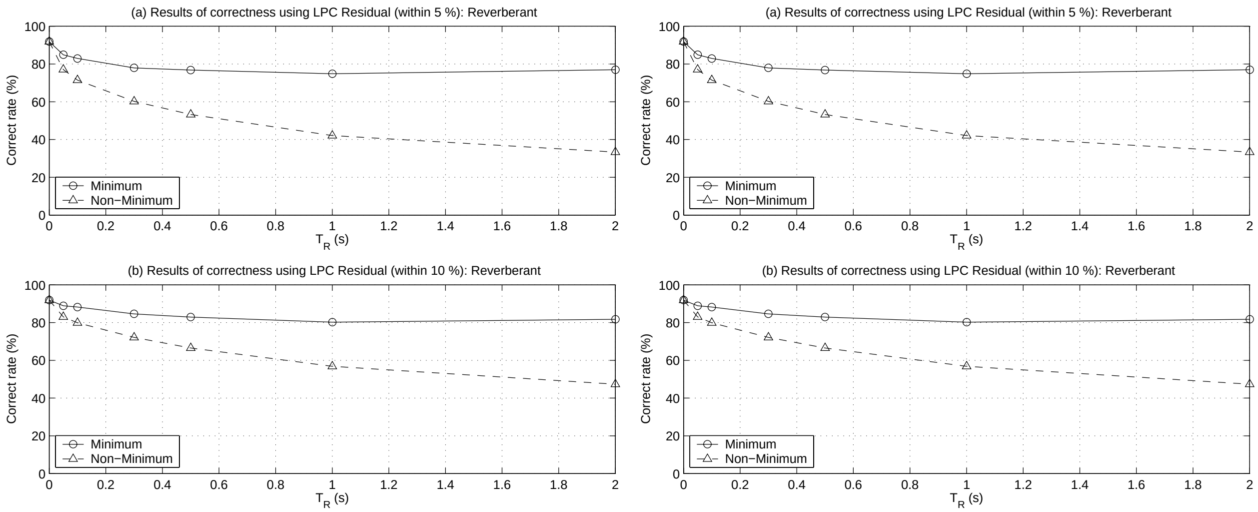

図 23 LPC 法 (Residu) による推定結果．(a) 5%誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差.

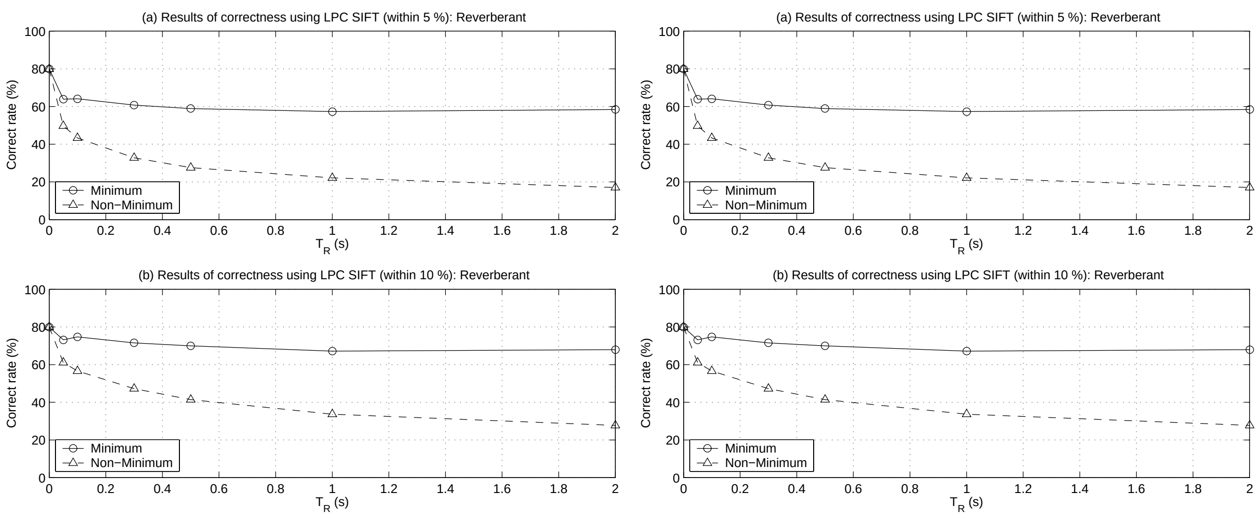

図 24 LPC 法 (SIFT 法) による推定結果．(a) 5%誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差.

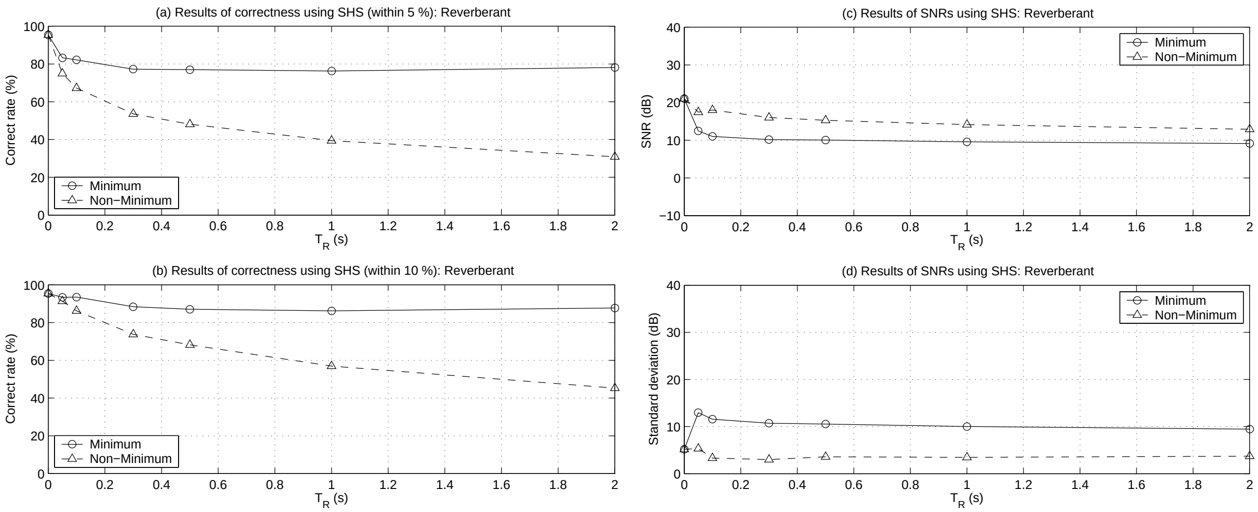

図 25　SHS 法による推定結果．(a) 5%誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差．

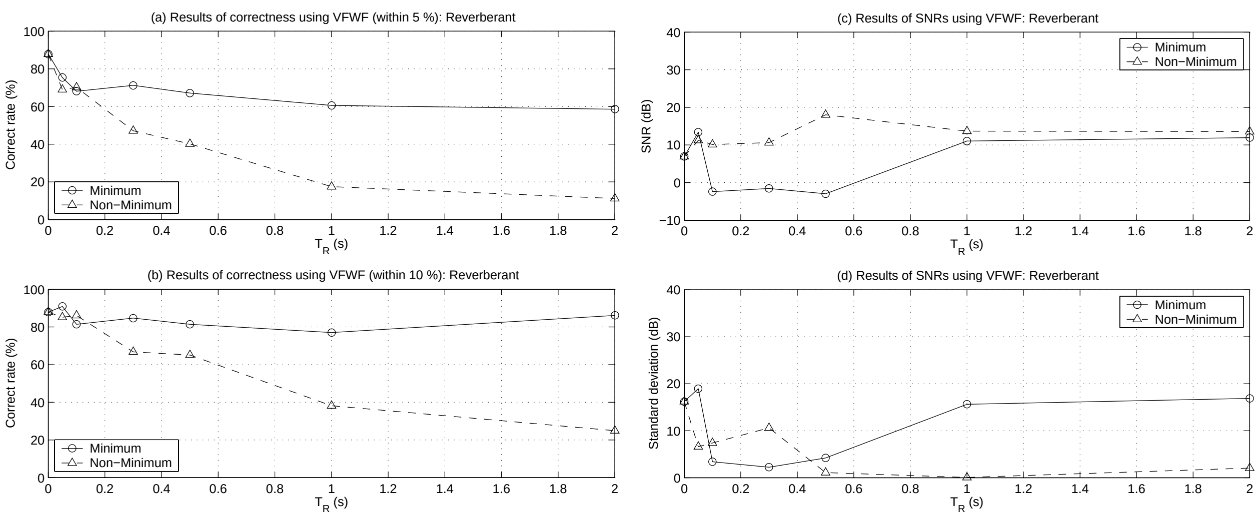

図 26　基本波フィルタ処理による推定結果．(a) 5%誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差．

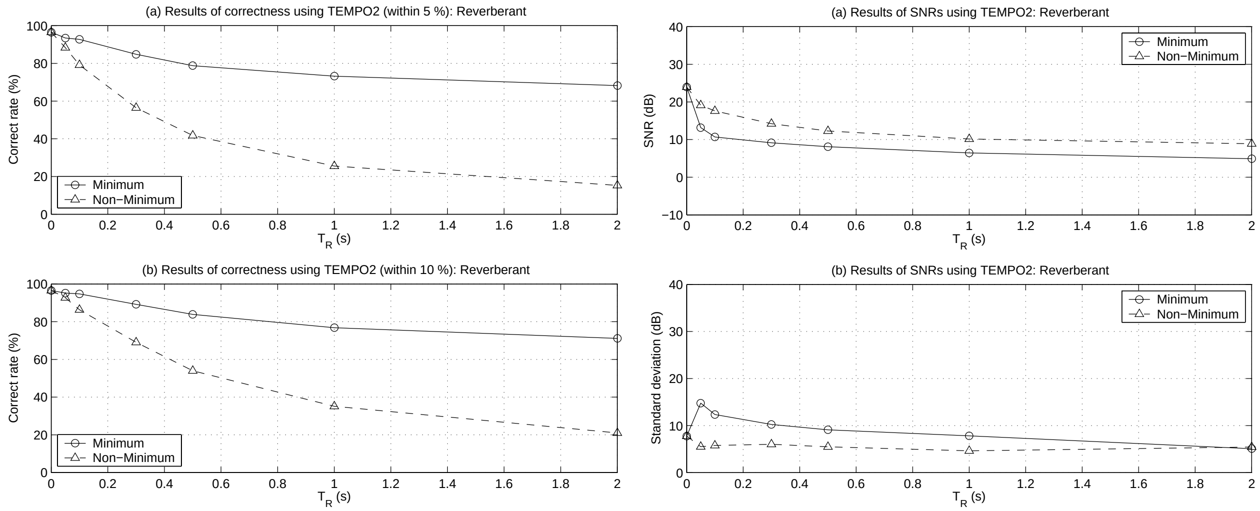

図 27　TEMPO による推定結果．(a) 5%誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差．

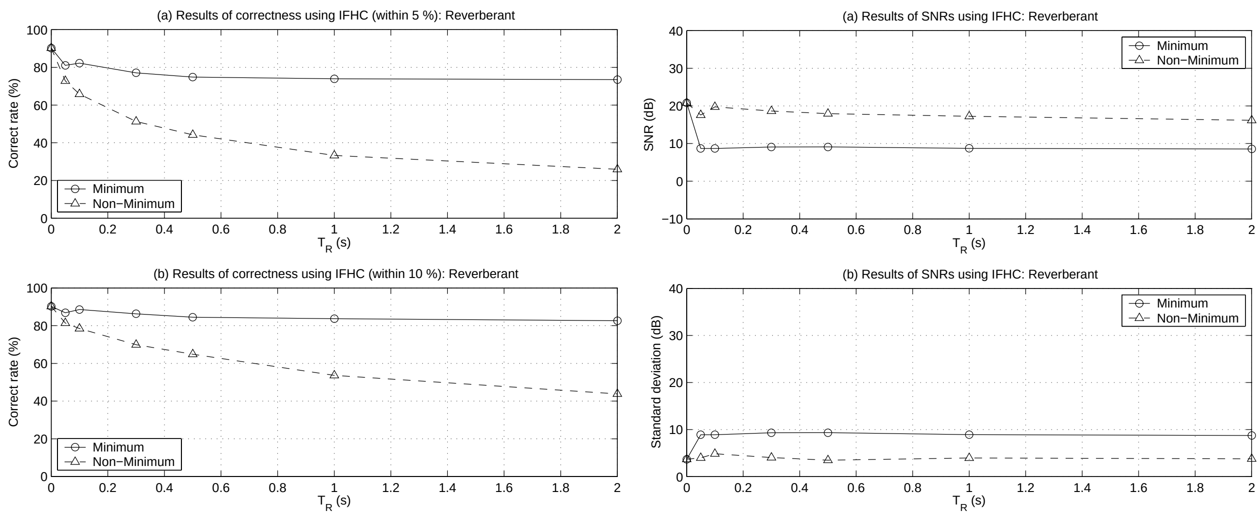

図 28　IFHC 法による推定結果．(a) 5%誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差．

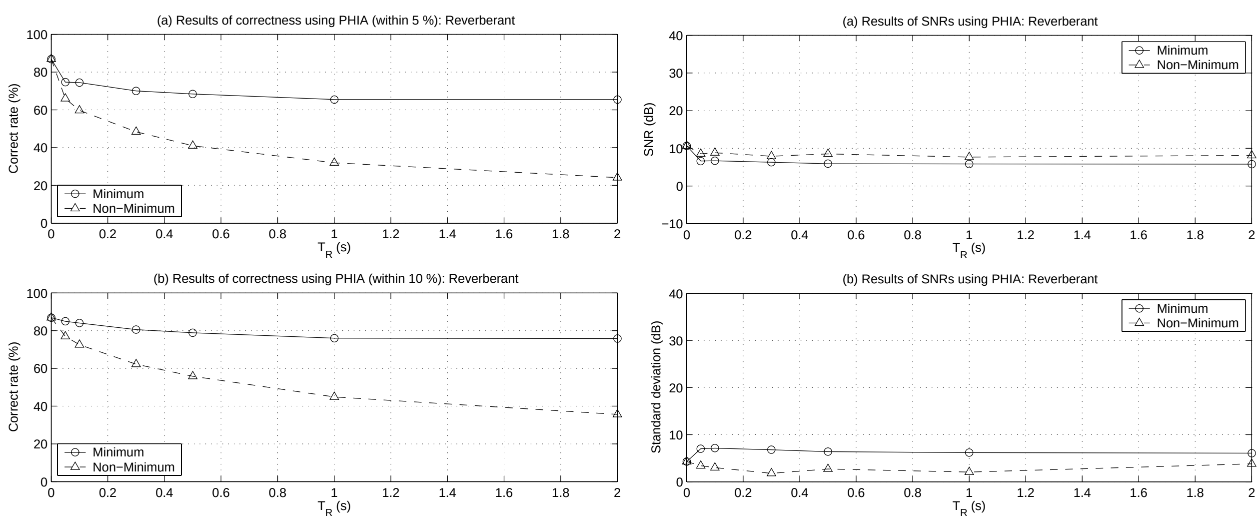

図 29　PHIA 法による推定結果．(a) 5%誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差．

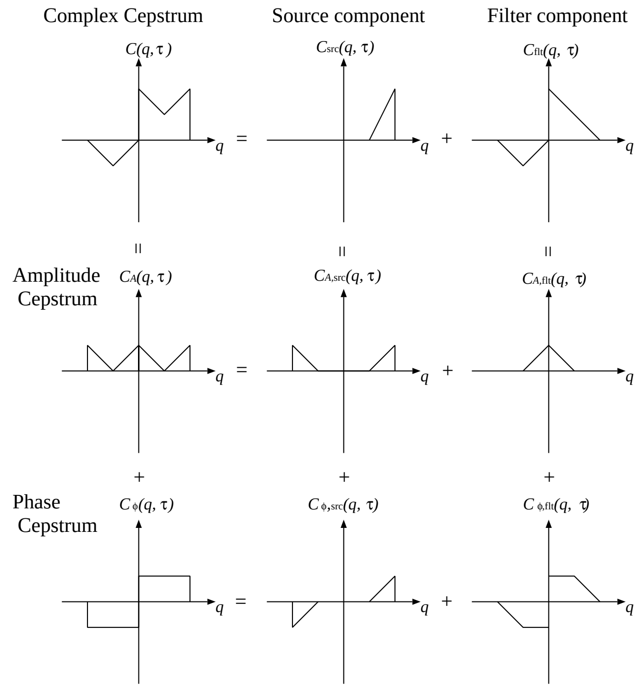

図 30　複素ケプストラム分析におけるソースフィルタモデルの関係.

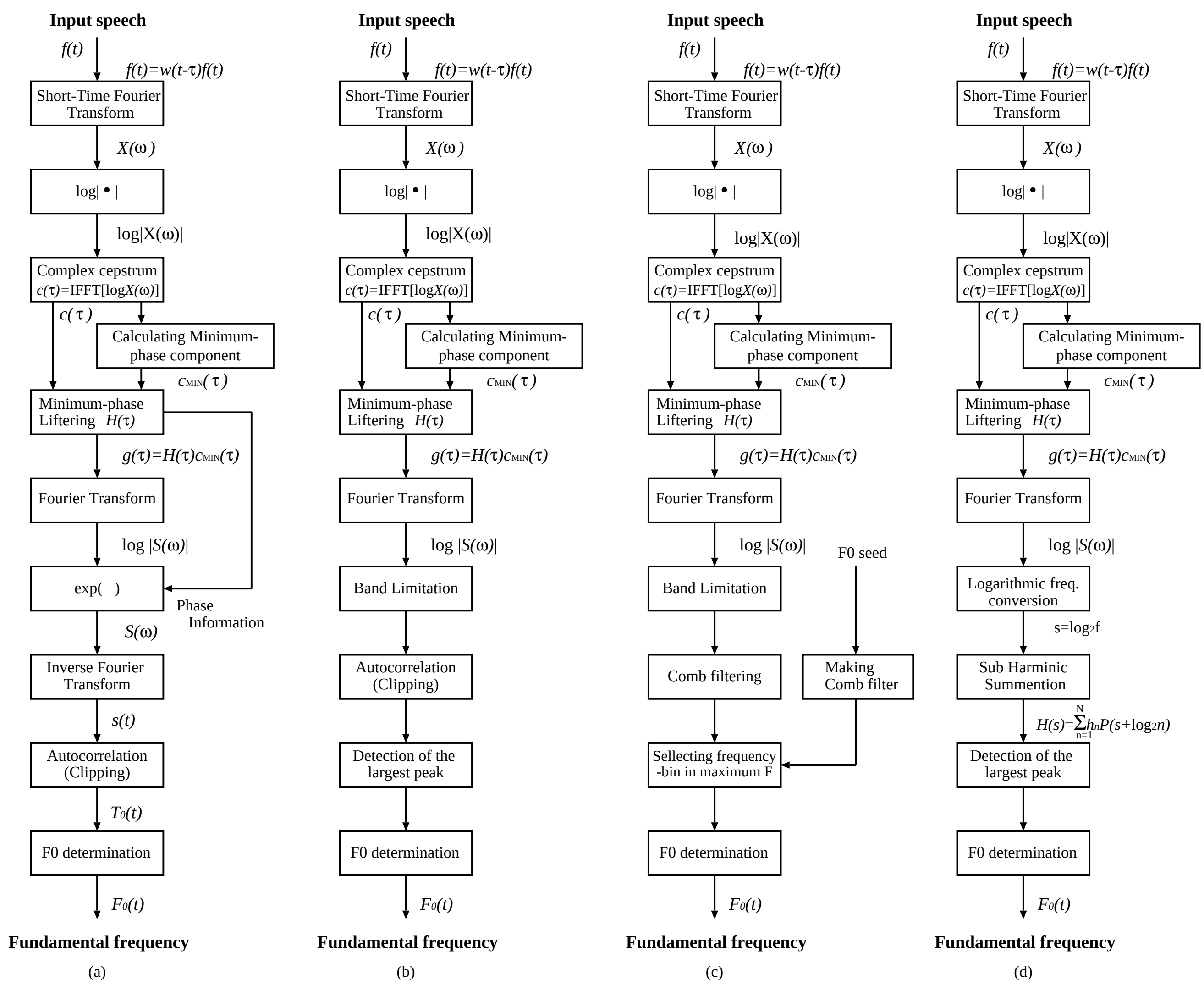

図 31　複素ケプストラム分析を利用したソースフィルタモデルに基づく基本周波数推定法のアルゴリズム．(a) 音源信号に対する自己相関法，(b) 音源スペクトルに対する自己相関法，(c) 音声スペクトルに対する Comb フィルタリング，(d) 音源スペクトルに対する SHS 法.

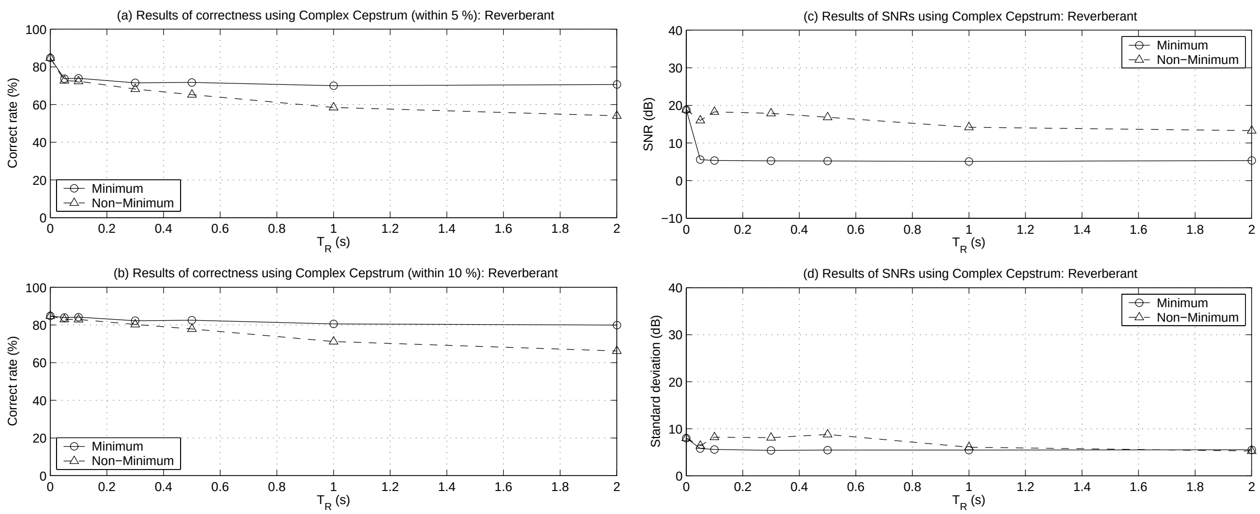

図 32　複素 Cepstrum 法（音源信号に対する自己相関法）による推定結果．(a) 5%誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差．

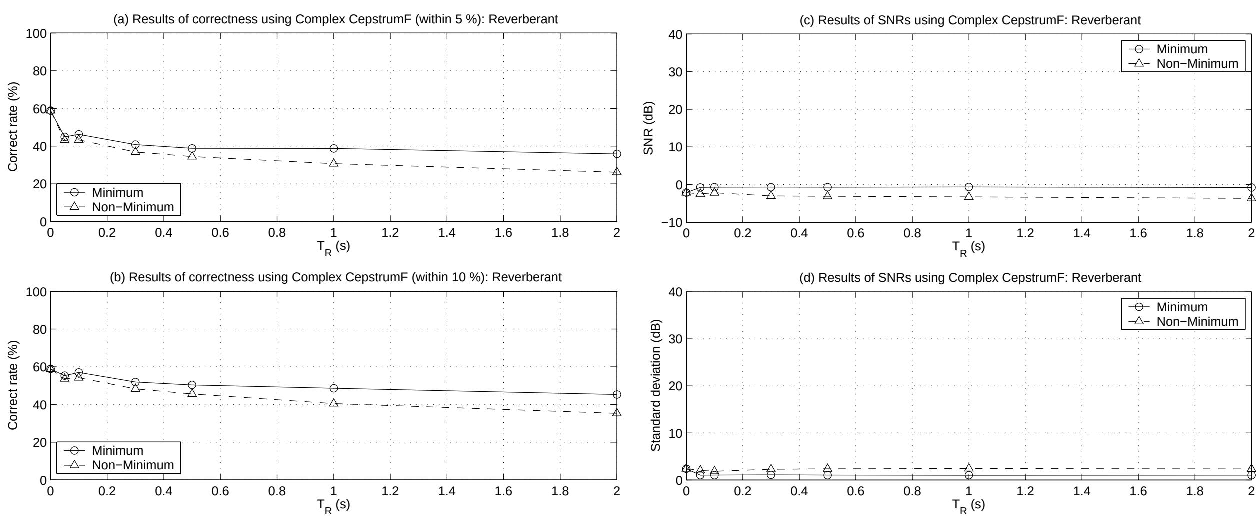

図 33　複素 Cepstrum 法（音源スペクトルに対する自己相関法）による推定結果．(a) 5%誤差内の正答率，(b) 10%誤差内の正答率，(c) 平均 SNR，(d) SNR の標準偏差．

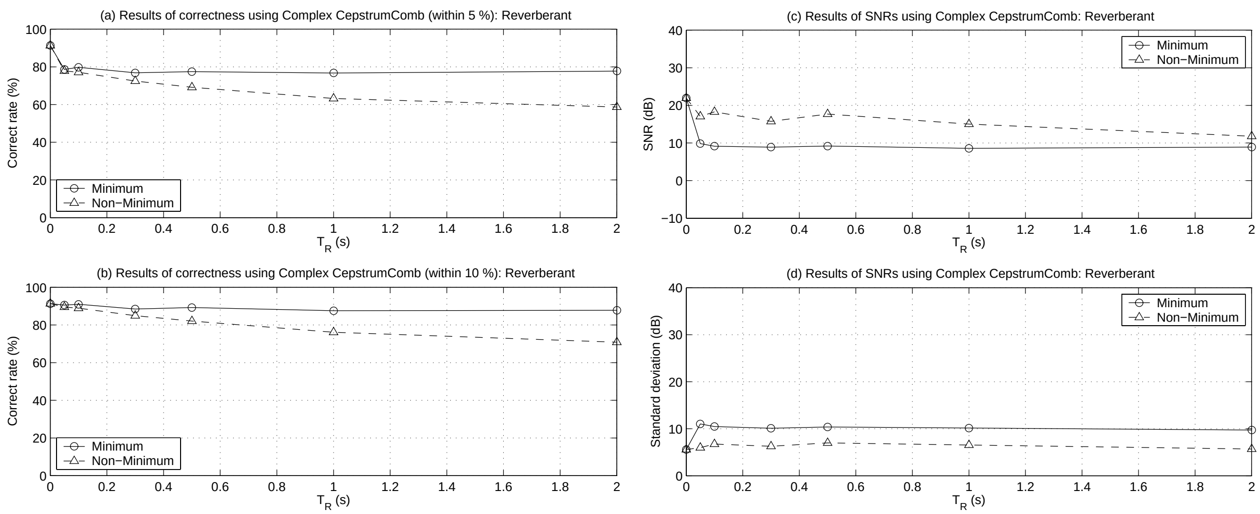

図 34　複素 Cepstrum 法（音源スペクトルに対する Comb フィルタリング）による推定結果.
(a) 5%誤差内の正答率,（b) 10%誤差内の正答率,（c) 平均 SNR,（d) SNR の標準偏差.

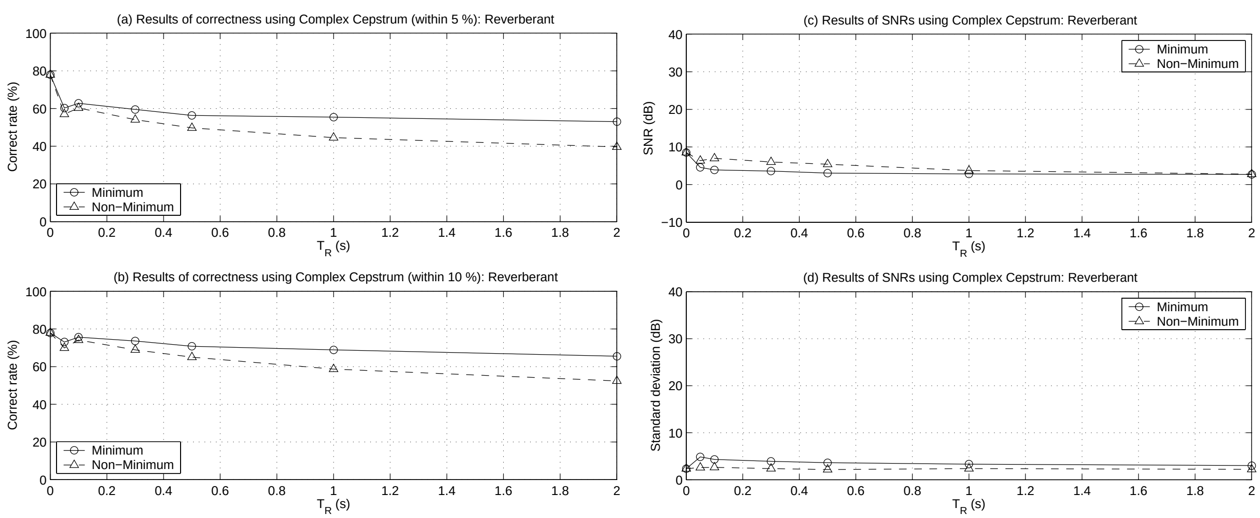

図 35　複素 Cepstrum 法（音源スペクトルに対する SHS 法）による推定結果.（a) 5%誤差内の正答率,（b) 10%誤差内の正答率,（c) 平均 SNR,（d) SNR の標準偏差.