論　文

対数スペクトルの任意基底関数による展開に基づく音声のスペクトル推定

若子　武士[†*]　　徳田　恵一[†]　　益子　貴史[††]　　小林　隆夫[††]
北村　　正[†]

Speech Spectral Estimation Based on Expansion of Log Spectrum by Arbitrary Basis Functions

Takeshi WAKAKO[†*], Keiichi TOKUDA[†], Takashi MASUKO[††], Takao KOBAYASHI[††], and Tadashi KITAMURA[†]

あらまし　メルケプストラムをパラメータとして音声スペクトルを表現した場合，1 次オールパス関数の m 段縦続接続を基底関数として対数スペクトルを表現することになる．本論文では，より効率的に音声スペクトルを表現するために，基底関数を任意の関数系から選ぶことのできるスペクトル推定法を提案し，そのための分析アルゴリズムを示す．本論文では特に，2 次のオールパス関数に基づいて定義された基底関数を用いる場合について考える．このとき，周波数目盛りの伸縮の度合に加えて，伸縮の中心となる周波数を設定することができるため，より自由度の高い周波数分解能の設定が可能となる．最後に，このスペクトル推定法を用いた分析合成音声の主観評価実験，及び HMM を用いた認識実験により，提案手法の有効性を示す．

キーワード　メルケプストラム，音声分析合成，音声認識，スペクトル推定

1. まえがき

有限個のメルケプストラム係数によって対数スペクトルを表現したとき，人間の聴覚特性と同様，低い周波数帯域に高い周波数分解能をもたせることができ，ケプストラムの半分近いパラメータ数で，聴覚的にはケプストラムによる場合と同等のスペクトル表現が可能である．このようなパラメータは音声認識で広く用いられているが [1], [2]，我々もこれまでに，メルケプストラムをパラメータとし，不偏ケプストラム法 [3] のスペクトル評価関数を最小化するように，メルケプストラムを決定するためのアルゴリズムを示し [4]（この手法は，音声信号をガウス過程と仮定した場合には最ゆう推定と等価となる），音声認識，音声符号化などでの有用性を示した [5]〜[7].

† 名古屋工業大学知能情報システム学科，名古屋市
Department of Computer Science, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya-shi, 466–8555 Japan
†† 東京工業大学大学院総合理工学研究科，横浜市
Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama-shi, 226–8502 Japan
* 現在，松下電器産業（株）AVC 社勤務

ところで，ケプストラム（厳密には複素ケプストラム）は，複素指数関数を基底関数として，対数スペクトルを展開した係数として与えられる．また，メルケプストラムは，同様に 1 次オールパス関数の m 段縦続接続を基底関数とした展開係数により与えられる．このことから，基底関数を任意の関数系から選ぶことにすれば，音声スペクトルにより適したスペクトル表現形式が得られる可能性がある．

本論文では，このような観点から，対数スペクトルを有限個の基底関数の線形結合によって表現したスペクトル推定法 [8] を考え，メルケプストラムをパラメータとする場合と同様の最小化問題が成立すること，及び最小解を得るためのアルゴリズムを示す．一例として，ウェーブレット基底により対数スペクトルを表現した場合のスペクトル推定例を示し，自由度の高いスペクトル表現が可能であることを示す．ただし，任意の基底関数から最適な基底関数を選ぶことは容易ではない．そこで，本論文では，2 次のオールパス関数により定義された基底関数を用いた場合 [9] について，音声分析合成，音声認識への応用を考える．この手法では，周波数目盛りの伸縮の度合に加えて，伸縮の中心

となる周波数を設定することができるため，自由度の高い周波数分解能の設定が可能となる．実際に，音声の分析合成及び音声認識に適用し，その有効性を示す．

2. 任意基底関数に基づいたスペクトル推定

2.1 任意基底関数に基づいたスペクトルモデル

対数振幅スペクトルを

$$\log\left|H(e^{j\omega})\right| = \sum_{m=0}^{M} c(m)\,\Psi_m(e^{j\omega}) \tag{1}$$

と，$M+1$ 個の線形独立な基底関数 $\Psi_m(z)$ によって表現することを考える．ただし，$\Psi_m(e^{j\omega})$ は実関数とする．$H(z)$ から，音声を再合成することを考えると，$H(z)$ は因果的で安定であることが望ましいため，任意の基底関数 $\Psi_m(e^{j\omega})$ に対して，$H(z)$ が最小位相となるように表現することにする．つまり，因果的で，その実部が $\Psi_m(e^{j\omega})$ に等しい関数 $\Phi_m(z)$（$\Phi_m(e^{j\omega})$ の虚部はその実部とヒルベルト変換により関係づけられる）を考える．$\Phi_m(z)$ は，

$$\phi_m(n) = \mathcal{Z}^{-1}\left[\Phi_m(z)\right] = 0, \quad n<0 \tag{2}$$

かつ，

$$\mathrm{Re}\left[\Phi_m(e^{j\omega})\right] = \Psi_m(e^{j\omega}) \tag{3}$$

の関係を満たし，$\Psi_m(z)$ の逆 z 変換を $\psi_m(n)$ とすれば，$\phi_m(n)$ は，

$$\phi_m(n) = \begin{cases} 2\,\psi_m(n), & n>0 \\ \psi_m(0), & n=0 \\ 0, & n<0 \end{cases} \tag{4}$$

で与えられる．このとき，

$$\log H(e^{j\omega}) = \sum_{m=0}^{M} c(m)\,\Phi_m(e^{j\omega}) \tag{5}$$

で与えられるシステム関数 $H(e^{j\omega})$ は，$\Phi_m(e^{j\omega})$ が因果的であることから最小位相であり，また，その対数振幅スペクトルは，式 (3) より，式 (1) で与えられる．

2.2 スペクトル評価関数

式 (1) あるいは式 (5) のスペクトルモデルに不偏ケプストラム法 [3] のスペクトル評価関数を適用する．

$$E(\boldsymbol{x},\boldsymbol{c}) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\exp R(\omega) - R(\omega) - 1\right\} d\omega \tag{6}$$

ただし，

$$R(\omega) = \log I_N(\omega) - \log\left|H(e^{j\omega})\right|^2 \tag{7}$$

$$\boldsymbol{x} = [x(0), x(1), \ldots, x(N-1)]^T \tag{8}$$

$$\boldsymbol{c} = [c(0), c(1), \ldots, c(M)]^T. \tag{9}$$

であり，T は転置を表している．また，$I_N(\omega)$ は，長さ N の時間窓 $w(n)$ を用いて得られた入力信号 $x(n)$ のピリオドグラム

$$I_N(\omega) = \left|\sum_{n=0}^{N-1} w(n)\,x(n)e^{-j\omega n}\right|^2 \Bigg/ \sum_{n=0}^{N-1} w^2(n) \tag{10}$$

である．式 (6) は，信号 $x(n)$ を弱定常過程としたとき，対数スペクトルの不偏推定量を得るための評価関数であるが，更に，信号 $x(n)$ が定常ガウス過程としたときには，この評価関数を最小化することは，漸近的に，そのゆう度関数 $P(\boldsymbol{x}\,|\,\boldsymbol{c})$ を最大化することに相当する [10].

2.3 最小化問題の解法

メルケプストラム分析法 [4] と同様，E は，$\boldsymbol{c}$ に関して下に（狭義の）凸であることが示される（付録参照）ので，大域的な一意解を，Newton-Raphson 法などの繰返し計算に基づいた最適化手法により容易に得ることができる．つまり，線形方程式

$$\boldsymbol{H}\,\Delta\boldsymbol{c}^{(i)} = -\nabla E\Big|_{\boldsymbol{c}=\boldsymbol{c}^{(i)}} \tag{11}$$

を解くことによって得られた $\Delta\boldsymbol{c}^{(i)}$ を用いて，

$$\boldsymbol{c}^{(i+1)} = \mathrm{c}^{(i)} + \Delta\boldsymbol{c}^{(i)}. \tag{12}$$

とする操作を E の変化が十分小さくなるまで繰り返す．ただし，

$$\Delta\boldsymbol{c}^{(i)} = [\Delta c^{(i)}(0), \Delta c^{(i)}(1), \ldots, \Delta c^{(i)}(M)]^T. \tag{13}$$

また，こう配 $\nabla E = \partial E/\partial\boldsymbol{c}$，及びヘッセ行列 $\boldsymbol{H} = \partial^2 E/\partial\boldsymbol{c}\,\partial\boldsymbol{c}^T$ は，

$$\boldsymbol{\Phi} = [\Phi_0(e^{j\omega}),\ \Phi_1(e^{j\omega}),\ \ldots,\ \Phi_M(e^{j\omega})]^T \tag{14}$$

とすれば，

$$\nabla E = -2 \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{I_N(\omega)}{|H(e^{j\omega})|^2} - 1 \right) \boldsymbol{\Phi}^* \, d\omega \tag{15}$$

$$\boldsymbol{H} = 2 \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|H(e^{j\omega})|^2} \boldsymbol{\Phi}^* (\boldsymbol{\Phi} + \boldsymbol{\Phi}^*)^T \, d\omega \tag{16}$$

で与えられる．通常の音声分析では繰返し回数は数回で十分であることがわかっている．

2.4 ウェーブレット基底によるスペクトル推定例

ここでは，$\phi_m(n) = \mathcal{Z}^{-1}[\Phi_m(z)]$ を離散ウェーブレット基底とした場合の結果を示す．$\phi_0(n) = \delta(n)$ とし，$\phi_m(n)$, $m = 1, 2, \ldots, M$ には，簡単にハール基底を考えた．周波数帯域分割は $0 \sim \pi/4$ (LL)，$\pi/4 \sim \pi/2$ (LH)，$\pi/2 \sim \pi$ (H) の 3 帯域とし，それぞれの帯域で用いた基底関数の数を M_{LL}, M_{LH}, M_H とする．それぞれの帯域の基底関数は，

$$\phi_{LL,m}(n) = \begin{cases} \dfrac{1}{2}, & 4m-3 \leqq n \leqq 4m \\ 0, & \text{otherwise} \end{cases}, \quad m = 1, 2, \ldots, M_{LL} \tag{17}$$

$$\phi_{LH,m}(n) = \begin{cases} \dfrac{1}{2}, & n = 4m-3, 4m-2 \\ -\dfrac{1}{2}, & n = 4m-1, 4m \\ 0, & \text{otherwise} \end{cases}, \quad m = 1, 2, \ldots, M_{LH} \tag{18}$$

$$\phi_{H,m}(n) = \begin{cases} \dfrac{\sqrt{2}}{2}, & n = 2m-1 \\ -\dfrac{\sqrt{2}}{2}, & n = 2m \\ 0, & \text{otherwise} \end{cases}, \quad m = 1, 2, \ldots, M_H \tag{19}$$

と書くことができ，式 (2)，(3) より，周波数領域では

$$\Psi_{LL,m}(e^{j\omega}) = \left(\cos\frac{\omega}{2} + \cos\left(\frac{3}{2}\omega\right) \right) \cos\left(\left(4m - \frac{3}{2}\right)\omega \right), \quad m = 1, 2, \ldots, M_{LL} \tag{20}$$

$$\Psi_{LH,m}(e^{j\omega}) = -\left(\sin\frac{\omega}{2} + \sin\left(\frac{3}{2}\omega\right) \right) \sin\left(\left(4m - \frac{3}{2}\right)\omega \right), \quad m = 1, 2, \ldots, M_{LH} \tag{21}$$

$$\Psi_{H,m}(e^{j\omega}) = -\sqrt{2} \sin\frac{\omega}{2} \sin\left(\left(2m - \frac{1}{2}\right)\omega \right), \quad m = 1, 2, \ldots, M_H \tag{22}$$

と書くことができる．ただし，$M = M_{LL} + M_{LH} + M_H$ である．男性話者による「人類学」という音声に対して，標本化周波数は 10 kHz とし，長さ 25.6 ms のブラックマン窓を用い，フレーム周期を 10 ms として分析を行った．図 1 にここで用いたウェーブレット基底のうちいくつかを示し，図 2 にいくつかの (M_{LL}, M_{LH}, M_H) の組合せに対するスペクトル推定

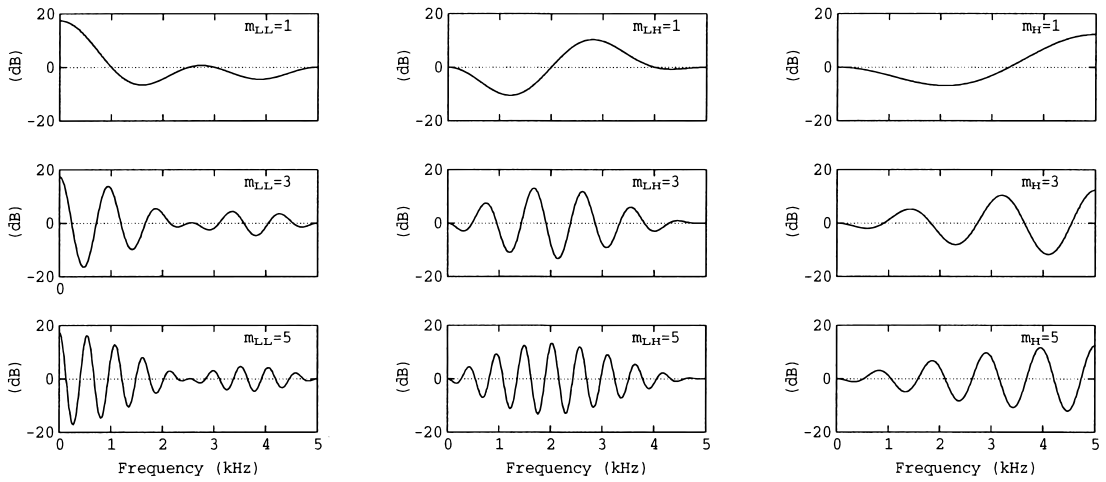

図 1　ウェーブレット基底例
Fig. 1　Wavelet basis functions.

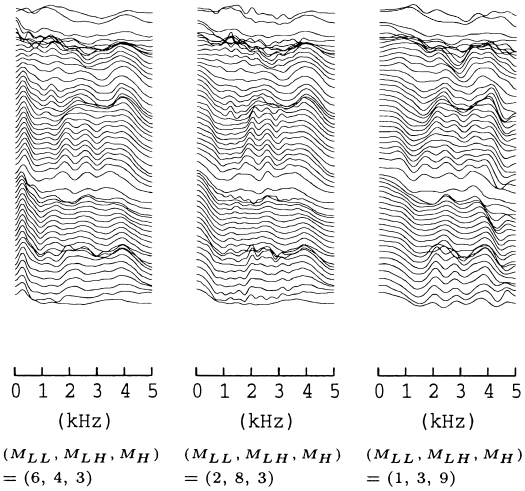

図 2　離散ウェーブレット基底を用いた推定例
Fig. 2　Spectral estimation by using wavelet basis functions.

例を示す．図より，係数を多く割り当てた帯域の周波数分解能が高くなっており，自由度の高い周波数分解能の設定が可能であることがわかる．

3. 2 次オールパス関数を利用したスペクトル推定法

前章では自由度の高いスペクトル表現が可能であることを例により示したが，任意の基底関数から最適な基底関数を選ぶことは容易ではない．そこで，本章以降では，2 次のオールパス関数により定義された基底関数を用いた場合について考えることにする．

3.1　2 次オールパス関数に基づいた基底関数

2 次オールパス関数を 1/2 乗したシステム関数

$$A(z) = \left(\frac{z^{-2} - 2\alpha\cos\theta z^{-1} + \alpha^2}{1 - 2\alpha\cos\theta z^{-1} + \alpha^2 z^{-2}} \right)^{\frac{1}{2}} \qquad (23)$$

を考える．このシステムの位相特性 $\tilde{\omega} = \beta(\omega)$ は，周波数 $0 \sim \pi$ を $0 \sim \pi$ に写像する単調な関数であり，周波数ワーピング関数として利用することができる．図 3 や図 4 (a) に例を示すように，α を 1 以下の正の値とすることにより，θ を中心とした周波数帯域を引き伸ばすことができる．ただし，θ は 2π (rad) を単位として表示している．このような 2 次オールパス関数の位相特性を周波数ワーピングに利用し，音声認識での有効性を調べたものに文献 [11] が挙げられる．しかし，この手法は，線形予測法によるスペクトル分析結果から周波数ワーピングされたケプストラムを求めるものであるため，2 次オールパス関数による周波数変換の特徴を生かしたスペクトル推定が行えるわけではない．

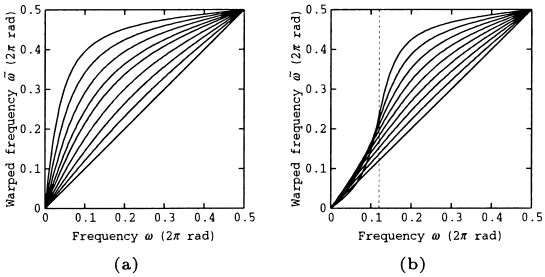

図 3　ワーピング関数 (a) $\theta = 0$, $\alpha = 0, \ldots, 0.8$ (b) $\theta = 0.12$, $\alpha = 0, \ldots, 0.8$
Fig. 3　Warping function. (a) $\theta = 0$, $\alpha = 0, \ldots, 0.8$ (b) $\theta = 0.12$, $\alpha = 0, \ldots, 0.8$.

ここでは，この 2 次オールパス関数を利用して，対数振幅スペクトルを

$$\begin{aligned} &\log\left|H(e^{j\omega})\right| \\ &\quad = \sum_{m=0}^{M} c(m)\,\cos(\tilde{\omega}m) \\ &\quad = \sum_{m=0}^{M} c(m)\,\left\{A^m(e^{j\omega}) + A^m(e^{-j\omega})\right\}/2 \end{aligned} \qquad (24)$$

つまり，式 (1) において

$$\Psi_m(e^{j\omega}) = \cos(\tilde{\omega}m) \qquad (25)$$

と表現し，式 (6) の評価関数を適用することを考える．前節と同様，$H(z)$ が最小位相となるように表現することを考えると，式 (4) より，$\Phi_m(z)$ は，

$$\phi_m(n) = \begin{cases} a_m(n) + a_m(-n), & n > 0 \\ a_m(0), & n = 0 \\ 0, & n < 0 \end{cases} \qquad (26)$$

で与えられる．ただし，

$$a_m(n) = \mathcal{Z}^{-1}\left[A^m(z)\right] \qquad (27)$$

である．$\theta = 0$ のときには，$A(z)$ は 1 次のオールパス関数，したがって，因果的な関数となるため，簡単に

$$\log H(e^{j\omega}) = \sum_{m=0}^{M} c(m)\,A^m(e^{j\omega}) \qquad (28)$$

と書くことができる．

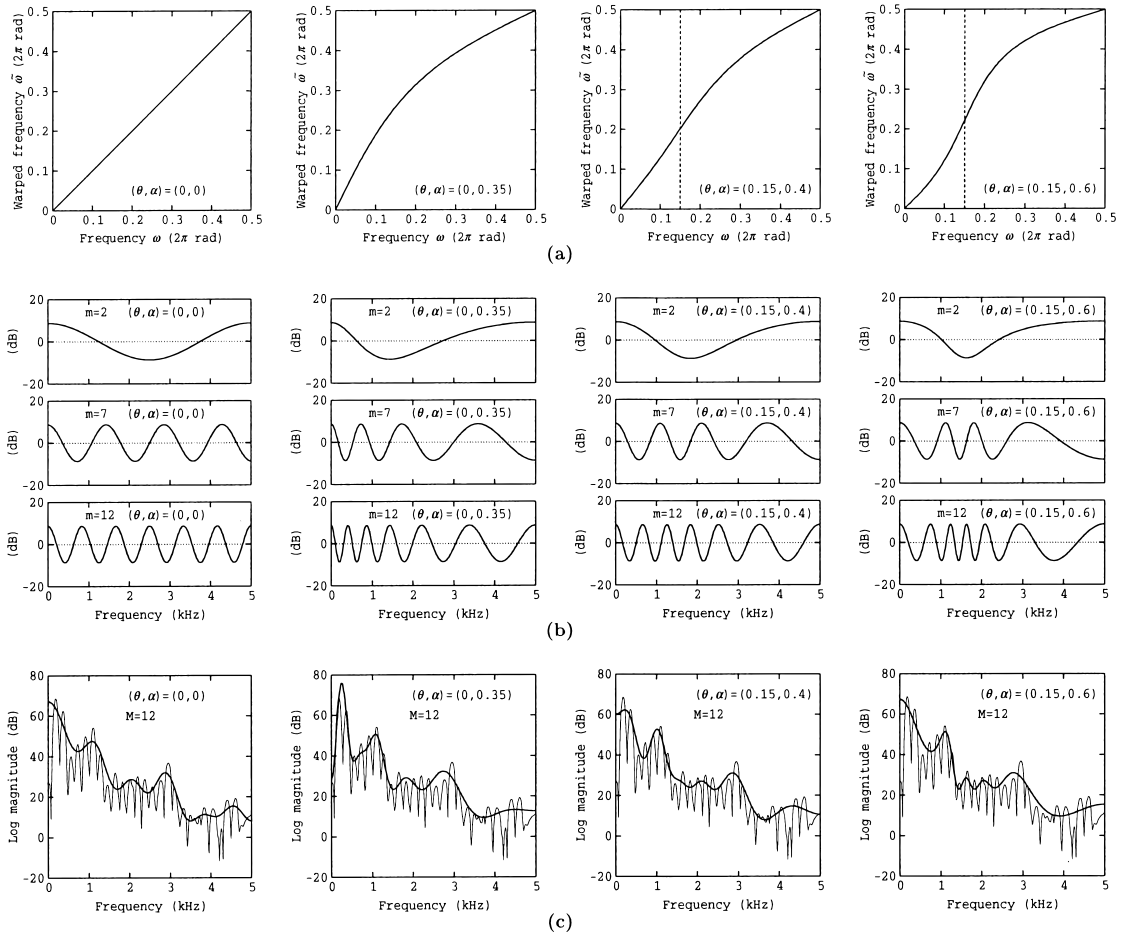

図 4　(a) ワーピング関数 (b) 基底関数 (c) スペクトル推定例
Fig. 4　(a) Warping function, (b) Basis function, (c) Spectral estimate.

基底関数 $\Phi_m(z)$ の対数振幅スペクトルを図 4 (b) に示す．$\alpha > 0$ の場合には，$(\theta, \alpha) = (0, 0)$ の基底を θ を中心に圧縮したものになっている様子がわかる．理論上，$\phi_m(n)$ は無限長の数列であるが，n の増加とともに急速に減衰するため，(θ, α) の値により，数十から 100 次程度までを用いれば十分な精度が得られる．

3.2　従来法との関係

$(\theta, \alpha) = (0, 0)$ とした場合には，

$$A^m(e^{j\omega}) = e^{-j\omega m} \tag{29}$$

となり，$c(m)$ は通常のケプストラム（厳密には最小位相の複素ケプストラム）となる．また，$\theta = 0$ とした場合には，

$$A^m(e^{j\omega}) = \left(\frac{e^{-j\omega} - \alpha}{1 - \alpha e^{-j\omega}}\right)^m$$

$$= e^{j\beta(\omega)m} \tag{30}$$

となり，$c(m)$ はメルケプストラムとなる．したがって，パラメータ θ, α を適切に選ぶことにより，ケプストラムやメルケプストラムによる場合に比べ，更に音声スペクトルの表現に適したスペクトル表現形式が得られる可能性がある．

3.3　スペクトル推定例

図 4 (c) にいくつかの (θ, α) の組合せに対するスペクトル推定例を示す．標本化周波数は 10 kHz とし，長さ 25.6 ms のブラックマン窓を用いた．$(\theta, \alpha) = (0, 0)$ の場合はケプストラムを，$(\theta, \alpha) = (0, 0.35)$ の場合は従来のメルケプストラムをパラメータとした分析法となる．図より，α の値を大きくするほど，周波数 θ 付近の帯域の周波数分解能が高くなっており，従来の

メルケプストラムによるスペクトル表現に対して，自由度の高い周波数分解能の設定が可能であることがわかる．

4. 応 用 例

4.1 音声分析例

電話帯域音声の符号化では，送話器特性を考慮した符号化を行う必要がある．ITU では送話器特性をシミュレートするためのフィルタ（IRS フィルタ）に関する勧告を行っており [12]，図 5 はその周波数特性である．図からわかるとおり，IRS 特性をもった音声スペクトルは周波数 0 付近の帯域が落ち込んだ特性となるため，従来のメルケプストラムを用いて推定されたスペクトルも周波数 0 付近の帯域が落ち込んだスペクトルとなり，この帯域のスペクトル形状を表現するために用いられた自由度は無駄になる．それに対して，2 次オールパス関数に基づく周波数変換を利用したメルケプストラム分析法を用いることにより，このような電話帯域音声の性質を考慮した分析が可能となる．

ここでは，標本化周波数 8 kHz の音声に対するスペクトル推定例を，IRS 特性を与えない場合，与えた場合について，それぞれ図 6 (a) 及び (b) に示す．θ の単位は 2π(rad) である．男性話者による「人類学」という音声に対し，長さ 32 ms のブラックマン窓を用い，フレーム周期は 10 ms として分析を行った．分析次数は 10 次である．$(\theta, \alpha) = (0.12, 0.6)$ の場合には，IRS 特性によって落ち込んだ周波数 0 付近の帯域を避けて，$\theta = 0.12$ 付近の帯域の周波数分解能を高くしている様子がわかる．

4.2 分析合成音声の主観評価試験

IRS 特性をもつ標本化周波数 8 kHz の音声を用いて，2 次オールパス関数に基づいたスペクトル分析による分析合成音の主観評価を行った．長さ 32 ms のブラックマン窓を用い，フレーム周期は 15.625 ms とした．分析次数は 10 次である．推定された最小位相システムのインパルス応答と励振源波形を畳み込むことにより，合成波形を得た．インパルス応答は 255 次までを用いた．テストデータには男性 2 名による文発声を用い，各話者から，それぞれ異なる 3 文章を選んだ．8 人の被験者に，原音声とともに，用意したデータセットを被験者ごとに異なるランダムな順に聞かせ，その品質の原音声に対する劣化の度合を 5 段階で評価させた（値が小さいほど劣化の度合が大きい）．受聴の繰返し回数は 1 回である．パラメータ (θ, α) は，予備試験によって選んだものを用いた．

試験により得られた DMOS（degradation mean opinion score）値を図 7 に示す．ここでは，θ の値ごとに結果の良かったものだけを載せている．図より，

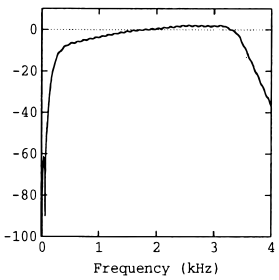

図 5 IRS 特 性
Fig. 5 IRS response.

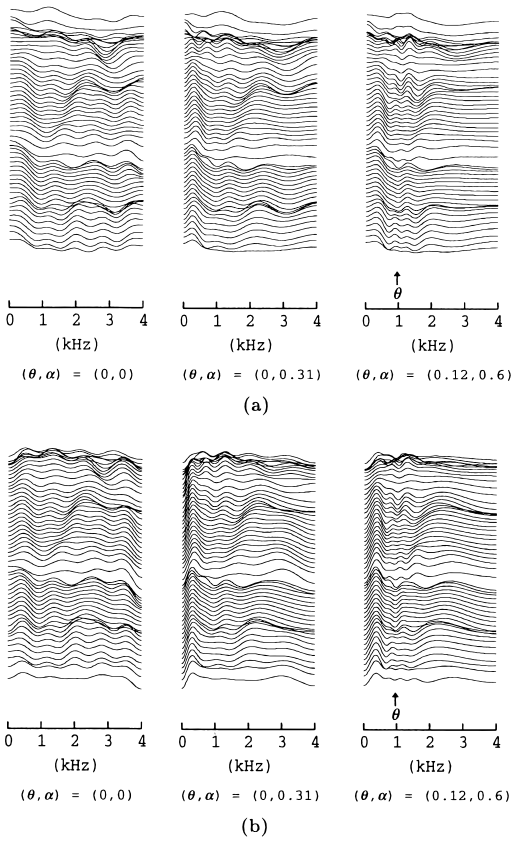

図 6 2 次オールパス関数を利用したスペクトル推定例 (a) IRS 特性をもたない場合 (b) IRS 特性をもつ場合
Fig. 6 Spectral estimation by using a second-order all-pass transfer function. (a) without IRS filter (b) with IRS filter.

$(\theta, \alpha) = (0.12, 0.6)$ の場合に最も良い評価を得ており，メルケプストラムに相当する $\theta = 0$ のときに最も結果の良い $\alpha = 0.26$ を用いた場合に対して，0.25 の MOS 値改善を，またケプストラムを用いた場合 $((\theta, \alpha) = (0, 0))$ に対して，1.06 の MOS 値改善を達成している．全体的に θ の値が大きいときの評価が高いのは，IRS 特性により落ち込んだ周波数 0 付近の帯域を避け，1 kHz 付近の周波数分解能を高くしているためと考えられる．ただし，1 kHz 以降の周波数分解能を高くした場合には，音声品質は急激に劣化したことを付記する．

なお，$(\alpha, \theta) = (0.0, 0.26)$ と $(0.12, 0.6)$ のスコアの有意差検定を行ったところ，危険率 20%，30%での有意なスコアの差はそれぞれ，0.29, 0.23 となったことから，0.25 のスコア差の危険率は 20%〜30%となる．音質的には，$(\alpha, \theta) = (0.0, 0.26)$ の場合に比べて $(0.12, 0.6)$ の場合の方が，音韻性がはっきりするように感じられたが，更に詳細な検討が必要と思われる．一方，$(\alpha, \theta) = (0.0, 0.00)$ と $(0.12, 0.6)$ のスコアの差 1.06 は，危険率 0.1%以下で有意となる．

4.3 音素認識実験

2 次オールパス関数に基づくスペクトル分析の音声認識における有効性を調べるため，IRS 特性をもつ標本化周波数 8 kHz の音声を用いて，HMM による音素認識実験を行った．実験は，音素に関する連続音声認識であり,「音素の列は音節の並びになる」という制約を課している．分析には，長さ 32 ms のブラックマン窓を用い，フレーム周期は 10 ms とした．分析次数は 12 次である．ATR 日本語音声データベースより，話者 MHT による 400 文章を学習データ，103 文章をテストデータとし，計 28 音素の平均正解精度を調べた．用いた HMM は 3 状態，混合数 2 のモノホンモデルである．なお，おおよそのパラメータ (θ, α) は，予備実験によって選出している．

実験結果を図 8〜図 10 に示す．図 8 は IRS 特性をもつ音声に対する，従来のメルケプストラム分析法を用いた場合の音素正解精度を，IRS 特性をもたない音声に対する結果と比較して示している．図より，IRS 特性をもつ音声の音素正解精度は，α の値にかかわらず，IRS 特性をもたない音声を下回っていることがわかる．これは，IRS 特性により周波数 0 付近の情報が欠落しているためと考えられる．

図 9 は，提案法において，θ の値を 0.10〜0.14 とした場合の IRS 特性をもつ音声の音素正解精度である．図より，各 θ の値における α の最適値は，θ の値の増加に伴って大きくなっているが，音素正解精度は $\theta = 0.12$ をピークとして徐々に低くなっている様子がわかる．ここでは $(\theta, \alpha) = (0.12, 0.31)$ のとき，すなわち，分析合成音声の主観評価実験結果と同様，960 Hz 付近の帯域を引き伸ばした場合に最も良い結果を得ている．図 10 は，IRS 特性の有無による，従来法と提案法における最適パラメータ時の音素正解精度の違いを比較したものである．IRS 特性をもたない音声に対しては，提案法において $(\theta, \alpha) = (0.04, 0.3)$ とした場合，すなわち 320 Hz 付近の帯域をを引き延ばした場合に最も高い正解精度を示したが，これは，$\theta = 0$ とした従来法で最も高い認識率を示した $(\theta, \alpha) = (0, 0.29)$ とほぼ同等の結果である．一方，IRS 特性をもつ音声

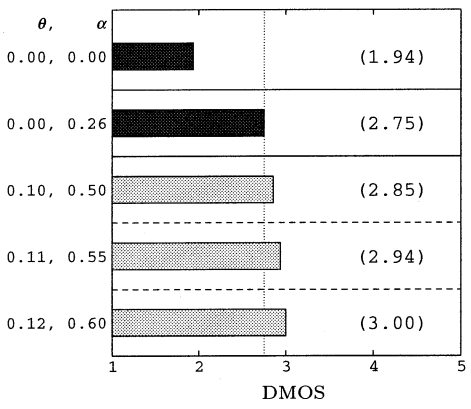

図 7　主観評価試験結果
Fig. 7　Result of a subjective performance test.

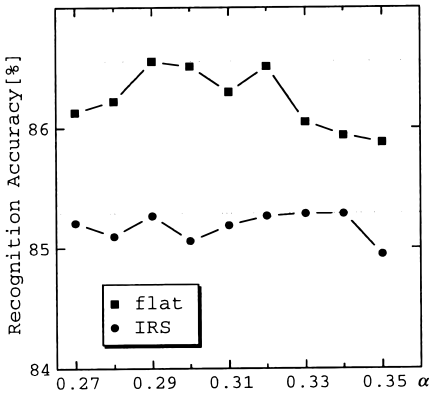

図 8　従来法による音素正解精度
Fig. 8　Recognition accuracy for the mel-cepstral analysis [2].

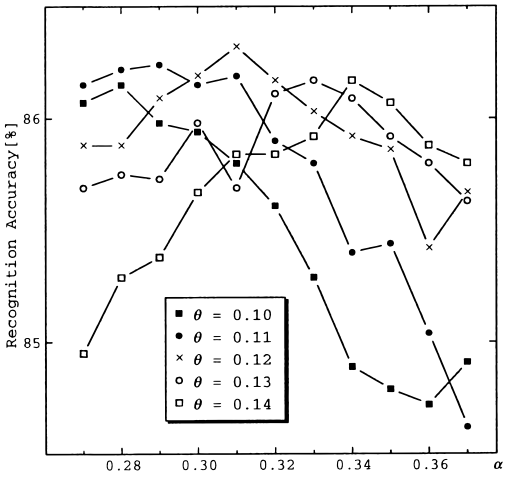

図 9　提案法による音素正解精度
Fig. 9　Recognition accuracy for the proposed technique.

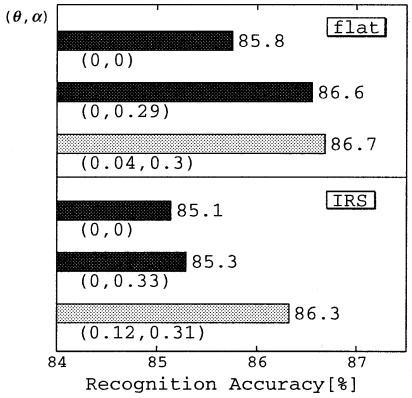

図 10　最適パラメータによる音素正解精度比較
Fig. 10　Comparison of recognition accuracy.

に対しては，提案法において，欠落した 0 付近の帯域を避けて引き伸ばすことにより，従来法を用いた場合に見られる音素正解精度の低下を抑えている．このとき，ケプストラム $((\theta, \alpha) = (0, 0))$ の結果に対して 7.9%，メルケプストラム $((\theta, \alpha) = (0, 0.33))$ の結果に対して 7.0%の認識正解精度改善率を達成している．

なお，実験結果の有意差に関して，3 種類の誤り（置換，挿入，削除）が総計されていること，音素ネットワークを使用していること，などの点から，厳密な検定を行うことは容易ではない．仮に，本実験をサンプル数 5000 個(注 1)，誤認識率 15%の単純な音素識別実験と考えた場合には，危険率 5%，10%，20%での有意な認識率の差はそれぞれ，1.14%，1.18%，0.91%となるが [13]，テストデータ数，実験方法，有意差検定法を含め，更に詳細な検討が必要である．

(注 1)：本実験のテストデータ中には 4759 個の音素があった．

5. むすび

対数スペクトルを任意の有限個の基底関数の線形結合によって表現したスペクトル推定法を提案した．特に，2 次オールパス関数により定義された基底関数を用いたスペクトル推定法を考え，信号の性質にあった θ, α の値を選ぶことによって効率的なスペクトル表現が可能であることを，IRS 特性をもった音声を例にとり，音声の分析合成実験と音素認識実験により示した．

今後の課題として，不特定話者による認識実験，θ, α の最適値の話者依存性に関する検討，他の性質をもつ音声への適用，話者認識への応用 [14] などが挙げられる．

謝辞　有益な御討論を頂いた千葉工業大学今井聖教授に感謝致します．また，ウェーブレット基底によるスペクトル分析プログラムを作成した一色直広氏（現キヤノン（株））に感謝致します．本研究の一部は，文部省科学研究費補助金萌芽的研究（課題番号 08875076）による．

文　献

[1] S.B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-33, pp.357–366, Aug. 1986.

[2] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," J. Acoust. Soc. America, vol.87, pp.1738–1752, April 1990.

[3] 今井　聖，古市千枝子，"対数スペクトルの不偏推定，" 信学論（A），vol.J70-A, no.3, pp.471–480, March 1987.

[4] 徳田恵一，小林隆夫，深田俊明，斎藤博徳，今井　聖，"メルケプストラムをパラメータとする音声のスペクトル推定，" 信学論（A），vol.J74-A, no.8, pp.1240–1248, Aug. 1991.

[5] K. Tokuda, T. Kobayashi, T. Masuko, and S. Imai, "Mel-generalized cepstral analysis — A unified approach to speech spectral estimation," Proc. of ICSLP-94, vol.3, pp.1043–1046, Sept. 1994.

[6] 徳田恵一，小林隆夫，深田俊明，今井　聖，"適応メルケプストラム分析を利用した音声の符号化とその評価，" 信学論（A），vol.J77-A, no.11, pp.1443–1452, Nov. 1994.

[7] K. Koishida, K. Tokuda, T. Kobayashi, and S. Imai, "CELP coding based on mel-cepstral analysis," Proc. of ICASSP, vol.1, pp.33–36, May 1995.

[8] 徳田恵一，益子貴史，小林隆夫，"対数スペクトルの任意

基底関数による展開に基づいた音声のスペクトル推定," 日本音響学会講論集, 1-P-22, pp.347–348, March 1997.
[9] 徳田恵一, 益子貴史, 小林隆夫, 北村　正, "2 次オールパス関数による周波数変換を利用した音声のメルケプストラム分析," 日本音響学会講論集, 3-7-13, March 1998.
[10] K. Dzhaparidze, Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series, Springer-Verlag, New York, 1986.
[11] 村上卓弥, 板倉文忠, "周波数ワープされたパラメータの音声認識における有効性の検討," 信学技報, SP93-17, June 1993.
[12] "Recommendation G.191 — Software tools for speech and audio coding standardization," ITU, 1993.
[13] 中川聖一, 高木英行, "パターン認識における有意差検定と音声認識システムの評価法," 日本音響学会誌, vol.50, no.10, pp.849–854, Oct. 1994.
[14] 宮島千代美, 渡辺秀行, 徳田恵一, 北村　正, 片桐　滋, "識別的特徴抽出に基づく話者認識—2 次オールパス関数に基づくメルケプストラム特徴の最適化," 日本音響学会講論集, 1-1-8, pp.35–36, March 1999.

付　　録

E が c に関して下に凸であることの証明

任意の異なる二つの $\boldsymbol{c}$ を $\boldsymbol{c}_a$, $\boldsymbol{c}_b$ として,

$$aE(\boldsymbol{c}_a)+(1-a)E(\boldsymbol{c}_b) > E(a\boldsymbol{c}_a+(1-a)\boldsymbol{c}_b), \quad 0<a<1 \qquad \text{(A·1)}$$

であれば, E は $\boldsymbol{c}$ に関して下に（狭義の）凸である. 式 (A·1) の左辺から右辺を引くと,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} I_N(\omega)\{af(x_a(\omega))+(1-a)f(x_b(\omega)) - f(ax_a(\omega)+(1-a)x_b(\omega))\}\,d\omega \qquad \text{(A·2)}$$

ただし,

$$f(x)=\exp(-2x) \qquad \text{(A·3)}$$

$$x_a(\omega)={\boldsymbol{c}_a}^T\boldsymbol{\Psi}(\omega) \qquad \text{(A·4)}$$

$$x_b(\omega)={\boldsymbol{c}_b}^T\boldsymbol{\Psi}(\omega) \qquad \text{(A·5)}$$

$$\boldsymbol{\Psi}(\omega)=[\Psi_0(e^{j\omega}),\ \Psi_1(e^{j\omega}),\ldots,\ \Psi_M(e^{j\omega})]^T \qquad \text{(A·6)}$$

と書くことができる. $f(x)$ は x に関して下に（狭義の）凸であることが容易に示される. したがって, (1) $I_N(\omega)\not\equiv 0$, (2) $\boldsymbol{c}_a\neq\boldsymbol{c}_b \Rightarrow {\boldsymbol{c}_a}^T\boldsymbol{\Psi}(\omega)\not\equiv{\boldsymbol{c}_b}^T\boldsymbol{\Psi}(\omega)$ が成り立てば, 式 (A·2) は正となり, E は $\boldsymbol{c}$ に関して下に（狭義の）凸であることが示される. 条件 (1) は通常の音声信号では常に成り立つと考えてよい. また, $\Psi_m(e^{j\omega})$ が線形独立ならば条件 (2) が成り立つ.

（平成 11 年 2 月 22 日受付, 5 月 31 日再受付）

若子　武士

平 9 名工大・工・知能情報システム卒. 平 11 同大大学院博士前期課程了（電気情報工学専攻）. 在学中, 音声分析の研究に従事. 現在, 松下電器産業（株）AVC 社勤務. 日本音響学会会員.

徳田　恵一　（正員）

昭 59 名工大・工・電子卒. 平 1 東工大大学院博士課程了. 同年東工大電気電子工学科助手. 平 8 名工大知能情報システム学科助教授. 工博. 音声分析, 音声合成・符号化, 音声認識, ディジタル信号処理の研究に従事. 日本音響学会, 人工知能学会, IEEE 各会員.

益子　貴史　（正員）

平 5 東工大・工・情工卒. 平 7 同大大学院博士前期課程了（知能科学専攻）. 同年同大大学院総合理工学研究科物理情報工学専攻助手. 音声の分析・合成, 音声認識の研究に従事. 日本音響学会, IEEE, ESCA 各会員.

小林　隆夫　（正員）

昭 52 東工大・工・電気卒. 昭 57 同大大学院博士課程了. 同年東工大精密工学研究所助手. 工博. 現在同大大学院総合理工学研究科物理情報工学専攻教授. ディジタルフィルタ, 音声の分析・合成, 音声認識の研究に従事. 日本音響学会, IEEE, ESCA 各会員.

北村　　正　（正員）

昭 48 名工大・工・電子卒. 昭 53 東工大大学院博士課程了. 同年東工大精密工学研究所助手. 昭 58 名工大・工・電子工学科講師. 昭 59 同助教授. 平 7 名工大知能情報システム学科教授. 工博. 音声情報処理, マルチメディア情報処理の研究に従事. 日本音響学会, IEEE, ESCA 各会員.