

スペクトル包絡抽出のための非一様スペクトル荷重 によるケプストラム分析

正員 徳田 恵一[†] 正員 小林 隆夫[†] 正員 今井 聖[†]

Cepstral Analysis with Non-Uniform Spectral Weighting for Spectral Envelope Extraction

Keiichi TOKUDA[†], Takao KOBAYASHI[†] and Satoshi IMAI[†], Members

あらまし 本論文では、対数振幅スペクトル上で非一様な荷重のスペクトル評価基準を用いることにより、スペクトル包絡を抽出するケプストラム分析法を提案している。従来のケプストラム法により得られる平滑化対数振幅スペクトルは、対数振幅スペクトルの線形平滑化による区間平均的な値をとるため、得られた平滑化対数スペクトルは、分析フレームごとにレベルが変動し不安定なものとなり、これが合成音声の品質や音声認識における認識率を低下させる原因となっていた。ところで、有声音においては、微細周波数構造のピークを、真のスペクトル包絡の標本点と考えることができる。そこで、本方法では微細周波数構造のピークにより大きな重みをかけることにより、微細周波数構造のピークを通るスペクトル包絡を得ている。本方法はピッチの抽出を必要としないため、有声音、無声音にかかわらず安定したスペクトル包絡を得ることができ、音声の自動分析に適している。具体例により本方法の有効性を示すとともに、他の手法、特に改良ケプストラム法との関係について検討している。

1. ま え が き

従来のケプストラム法⁽¹⁾により得られる平滑化対数振幅スペクトルは、対数振幅スペクトルの線形平滑化による区間平均的な値をとるため、特に有声音において、窓関数による微細周波数構造の谷部の乱れに大きく影響を受ける。従って、得られた平滑化対数スペクトルは、分析フレームごとにレベルが変動し、不安定なものとなっており、これが合成音声の品質や、音声認識における認識率を低下させる原因となっていた。この問題を避けるためには、微細周波数構造のピークを通る対数スペクトル包絡を抽出する必要がある。

そのための方法として、微細周波数構造のピークを探索し、直線で結ぶ方法がいくつか提案されている^{(2),(3)}。しかし、これらの手法は、ピッチ周期の抽出が必要のため、基本的には有声音に対してのみ有効であり、また誤ったピッチを抽出した場合の悪影響が大きい。

文献(4)は、窓関数の影響をあらかじめ考慮したケブ

ストラム法であるが、この場合もやはりピッチの抽出を必要とし、有声音に対してのみ有効である。また、文献(5)は、信号のスペクトルと合成されたスペクトルとのマッチングをとる方法で、誤差を最小にするピッチを探索することによりスペクトル包絡を得るが、ピッチ探索に手間がかかり、計算量を減らすためには正確なピッチの推定値が必要となる。

音声認識、残差あるいはマルチパルスなどを音源とする分析合成、規則合成のためのファイル作成などの音声分析においては、ピッチ情報は不要なことが多く、スペクトル分析はピッチ抽出なしで行えた方が都合が良いと考えられる。

そこで、本論文では微細周波数構造のピーク探索あるいはピッチ抽出などを必要としない、ケプストラム法に基づいた対数スペクトル包絡の抽出法を提案している。本方法は、対数振幅スペクトル上で非一様なスペクトル荷重を用いることにより、微細周波数構造のピークを通る対数スペクトル包絡を得るものである。本方法では、微細周波数構造のピーク探索あるいはピッチ抽出などが不要なため、有声音、無声音にかかわらず安定した包絡を得ることができ、繰り返し計算の

[†] 東京工業大学精密工学研究所、横浜市
Research Laboratory of Precision Machinery and Electronics,
Tokyo Institute of Technology, Yokohama-shi, 227 Japan

収束性も保証されている。

本方法により、微細周波数構造のピークを通るといいう意味で良好な対数スペクトル包絡が得られることを具体例で示し、得られた包絡は分析フレーム方向にもよく安定していることを示す。

微細周波数構造のピーク探索あるいはピッチ抽出などを行わず、微細周波数構造のピークを通る対数スペクトル包絡を抽出する方法として他に改良ケプストラム法⁽⁶⁾が提案されている。この方法は、平滑化操作の繰り返しにより、微細周波数構造のピークを通るスペクトル包絡を得るものである。本論文では、改良ケプストラム法との関係についてスペクトル荷重の観点から検討しており、その他の手法との関係についても簡単に考察している。

2. 対数スペクトル包絡の抽出法

2.1 ケプストラム法⁽¹⁾

音声信号は、短い区間について考えれば、線形時不変システムつまり声道を周期的なインパルス列または白色雑音で励振することにより生成されると考えることができる。このとき、音声信号 $x(n)$ は、

$$x(n)=h(n)*e(n) \tag{1}$$

のように音源波形 $e(n)$ と声道のインパルス応答 $h(n)$ の畳み込みで表される。ここで、記号 $*$ は畳み込みを表す。

$x(n)$ が安定ならば、フーリエ変換 $X(\omega)$ が存在し、

$$X(\omega)=\sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n} \tag{2}$$

と表され、 $X(\omega)$ の対数振幅を取った

$$X_R(\omega)=\ln|X(\omega)| \tag{3}$$

を対数振幅スペクトルと呼ぶ。ケプストラム c_m は、対数振幅スペクトル $X_R(\omega)$ の逆フーリエ変換で定義され、次の関係を持つ。

$$c_m=\frac{1}{2\pi}\int_{-\pi}^{\pi} X_R(\omega)e^{jm\omega}d\omega \tag{4}$$

$$X_R(\omega)=\sum_{m=-\infty}^{\infty} c_me^{-jm\omega}, (c_m=c_{-m}) \tag{5}$$

時間領域で畳み込まれた音源波形 $e(n)$ と声道のインパルス応答 $h(n)$ は、ケフレンシ領域では、 $e(n)$ のケプストラムを c_m^e 、 $h(n)$ のケプストラムを c_m^h として

$$c_m=c_m^e+c_m^h \tag{6}$$

と加法の関係に変換される。従って、それぞれのケプストラムに重なりがなければ、二つの信号成分を分離することができる。インパルス応答 $h(n)$ はケプストラムの低次係数に、音源波形 $e(n)$ はケプストラムの

高次係数に現れるため、低次の係数のみを取り出してフーリエ変換することにより、声道のインパルス応答 $h(n)$ の対数振幅特性は、平滑化対数振幅スペクトルとして

$$S(\omega)=\sum_{m=-M}^Mc_me^{-jm\omega} \tag{7}$$

で得られる。

2.2 ケプストラム法の問題点⁽⁷⁾

実際に音声信号を分析する際には、窓 $w(n)$ をかけるため、信号 $x(n)$ は、

$$x(n)=(h(n)*e(n))\cdot w(n) \tag{8}$$

の形になる。窓 $w(n)$ がインパルス応答 $h(n)$ の実効長に比べて、十分長いとすれば、

$$x(n)\simeq h(n)*(e(n)\cdot w(n)) \tag{9}$$

となり、ケプストラム法により、 $h(n)$ と $e(n)\cdot w(n)$ を分離することができる。しかし、実際にはこの仮定は成り立っておらず、スペクトルは、元のスペクトルに窓関数のスペクトルが畳み込まれたものとなり、正しく $h(n)$ を分離することはできない。このような窓の影響は無声音ではさほど大きくないが、有声音では線スペクトルに窓関数のスペクトルが畳み込まれることになるため、微細周波数構造の谷部に著しい影響が現れる。

ここでは、 $x(n)$ が周期 N_p の周期信号つまり有声音である場合について考えることにし、音源波形 $e(n)$ が

$$e(n)=\sum_{k=-\infty}^{\infty} N_p\cdot\delta(n-k\cdot N_p) \tag{10}$$

の形をしているとする。ただし、 $\delta(n)$ は単位インパルスを表す。このとき、式(8)のフーリエ変換は、声道のインパルス応答 $h(n)$ のフーリエ変換を $H(\omega)$ 、窓のフーリエ変換を $W(\omega)$ として、

$$X(\omega)=\sum_{k=-N_p/2}^{N_p/2} H(k\cdot\omega_p)\cdot w(\omega-k\cdot\omega_p) \tag{11}$$

と表される。ただし、 $\omega_p=2\pi/N_p$ である。窓 $w(n)$ が、

$$W(0)=\sum_{k=-\infty}^{\infty} W(k)=1 \tag{12}$$

と正規化されているとすれば、以下の仮定により微細周波数構造のピークつまり $\omega=k\cdot\omega_p$ における $X(\omega)$ は、元のスペクトル $H(\omega)$ と同じ値を持つと考えることができる。

(1) 窓関数のスペクトルのサイドローブが、隣の微細周波数構造のピークで、十分減衰している。

(2) 実際の計算では周波数軸は離散化されているが、窓関数スペクトル $W(\omega)$ のピークの平坦さに比

べ、周波数上での標本点は十分細かく、微細周波数構造のピーク付近の値を標本点としている。

従って、有声音においては、微細周波数構造のピークを通る対数スペクトル包絡をとることにより、声道の伝達特性 $H(\omega)$ の対数振幅特性が得られると考えられる。

微細周波数構造の谷部は、ピッチの変化、背景雑音、あるいは変化が急な谷部を標準化することなどによっても乱れる。従って、このような微細周波数構造の谷部の影響を避けるためにも微細周波数構造のピークを通るスペクトル包絡をとる必要がある。

2.3 ケプストラム法の改良法

改良ケプストラム法⁽⁶⁾では、平滑化対数振幅スペクトル $S(\omega)$ より上の部分にある対数振幅スペクトルを、更に平滑化するという操作を繰り返すことにより、対数振幅スペクトルの微細構造のピークを通る包絡を得ている。

求めようとする対数スペクトル包絡を

$$V(\omega) = \sum_{m=-M}^M u_m e^{-j\omega m}, \quad (u_m = u_{-m}) \quad (13)$$

とし、対数振幅スペクトル $X_R(\omega)$ に対し、残差スペクトル $R(\omega)$ を次のように定義する。

$$R(\omega) = X_R(\omega) - V(\omega) \quad (14)$$

i 番目の近似包絡 $V^{(i)}(\omega)$ に対し、

$$r(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (15)$$

として、 $r(R(\omega))$ を平滑化したものを加えることにより、次の近似値 $V^{(i+1)}(\omega)$ を得る。

この操作をケプレンシ領域で表すと以下になる。つまり、 i 番目の近似 $u^{(i)}_m$ に対し

$$\Delta u_m = (1 + \alpha) \frac{1}{2\pi} \int_{-\pi}^{\pi} r(R(\omega)) e^{j\omega m} d\omega \quad (16)$$

を計算し、次の近似値

$$u^{(i+1)}_m = u^{(i)}_m + \Delta u_m, \quad (m = -M, \dots, -1, 0, 1, \dots, M) \quad (17)$$

を得る。ここでは、 u_m の初期値 $u^{(0)}_m$ として平滑化対数振幅スペクトルのケプストラム c_m を用いることにする。

α は加速係数で、繰り返しの収束を速める働きをするが、その値を大きくとり過ぎると包絡が微細周波数構造のピークを越えることがあるので、実験的に得られた最適値を用いる必要がある。

3. 非一様なスペクトル荷重による対数スペクトル包絡の抽出

3.1 非一様なスペクトル荷重

式(13)、(14)と同様、対数振幅スペクトル $X_R(\omega)$ と対数スペクトル包絡 $V(\omega)$ との残差 $R(\omega)$ を

$$R(\omega) = X_R(\omega) - V(\omega) \quad (18)$$

とする。従来のケプストラム法では、2乗平均誤差

$$\varepsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} R^2(\omega) / 2 d\omega \quad (19)$$

を最小化する u_m を求めていることになり、このときのスペクトル荷重は、図1(a)のようになる。対数スペクトル包絡は、対数振幅スペクトル $X_R(\omega)$ を線形平滑化したものとなり、微細周波数構造の影響を受ける。

本方法では、微細周波数構造のピークを通る包絡を求めるため、式(19)を変形し、

$$\varepsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(R(\omega)) \cdot R^2(\omega) / 2 d\omega \quad (20)$$

ただし、

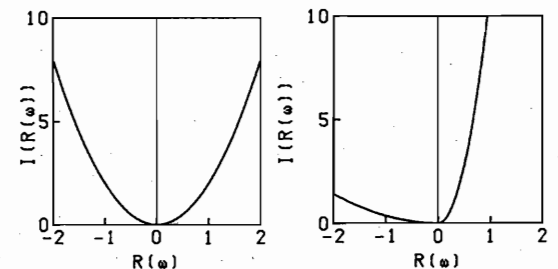
$$f(x) = \begin{cases} 1/\sigma_1^2, & x < 0 \\ 1/\sigma_2^2, & x \geq 0, \end{cases} \quad (21)$$

のスペクトル評価を考えることにする。 $\sigma_1^2 > \sigma_2^2$ とすれば、正の誤差が負の誤差に比べ厳しく評価され、包絡は微細周波数構造のピークに近い方に移動し、 $\sigma_1^2 \gg \sigma_2^2$ とすることにより、包絡は微細周波数構造のピークを通ると考えられる。ここで、 σ_1^2, σ_2^2 の値はその比だけが意味を持つので、

$$K = \sigma_1^2 / \sigma_2^2 \quad (22)$$

とおくことにする。図1(b)には、式(20)において $K = 32$ とした場合のスペクトル荷重を示してある。

式(20)の ε を最小にする u_m を求めるためには、式(20)を u_m で微分し零とおいて得られる



(a) Cepstral method $I(R(\omega)) = R^2(\omega)/2$ (b) Proposed method $I(R(\omega)) = f(R(\omega)) \cdot R^2(\omega)/2$

図1 スペクトル荷重

Fig. 1 Curves of the spectral weightings.

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} f(R(\omega)) \cdot R(\omega) e^{jk\omega} d\omega = 0$$
$$(k = -M, \dots, -1, 0, 1, \dots, M) \quad (23)$$

を解けばよい。

3.2 スペクトル残差の確率分布

式(20)の誤差評価は、スペクトル残差の確率分布が、次式のように $R(\omega)$ が負の部分で σ_1^2 、正の部分で σ_2^2 の分散を持つ非対称な正規分布であると仮定し、対数振幅スペクトル上の各点に対して最尤う推定法を適用したことに相当する。

$$P(R(\omega)) = \begin{cases} \frac{1}{(2\pi)^{1/2} \sigma_1} \exp\left(-\frac{R^2(\omega)}{2\sigma_1^2}\right), & R(\omega) < 0 \\ \frac{1}{(2\pi)^{1/2} \sigma_2} \exp\left(-\frac{R^2(\omega)}{2\sigma_2^2}\right), & R(\omega) \geq 0 \end{cases} \quad (24)$$

ただし、

$$\sigma = (\sigma_1 + \sigma_2)/2. \quad (25)$$

なぜならば、このときの u_m に対する対数ゆう度は、

$$\log L(R(\omega)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log P(R(\omega)) d\omega \quad (26)$$

であり、この対数ゆう度を最大にする u_m を求めるため、 u_m で微分し零とおくと、式(23)と同じものが得られる。

スペクトル残差 $R(\omega)$ が実際にどのように分布しているかを調べた1例を図2に示す。信号は、有声音の声道特性をシミュレートすると考えられるデジタルフィルタを、周期 10 ms のインパルス列で励振したものを、真の対数スペクトル包絡つまりデジタルフィルタの対数振幅特性とフィルタ出力の対数振幅スペクトルとの差を、43 フレームにわたりヒストグラムにしている。フィルタの共振点の中心周波数 (Hz)/帯域幅 (Hz) は、700/130, 1220/70, 2600/160 とした。標本化周波数は 10 kHz、時間窓は 25.6 ms のブラックマン窓を移動周期 6.4 ms で用い、フーリエ変換は

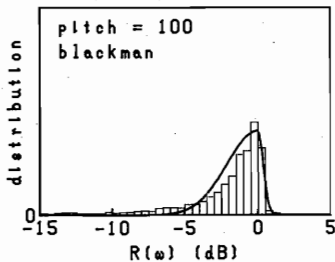


図2 スペクトル残差 $R(\omega)$ のヒストグラムと確率分布
Fig. 2 Plots of the measured and theoretical histogram of the log spectral residue $R(\omega)$.

256 点の FFT により計算した。

図には、 $K=32$ の場合について式(24)の $P(R(\omega))$ を重ねて描いてあるが、実際の分布ともよく合っており、このような確率分布の仮定が妥当なものであることがわかる。

3.3 解 法

式(23)は非線形方程式なので、Newton-Raphson 法を用いる。つまり、式(23)の左辺を $g_k(k = -M, \dots, -1, 0, 1, \dots, M)$ とおき、 i 番目の近似値 $u^{(i)}_m$ に対し、

$$\sum_{l=-M}^M \frac{\partial g_k}{\partial u_l} \Delta u_l = -g_k \quad (27)$$

を Δu_m について解き、次の近似値

$$u^{(i+1)}_m = u^{(i)}_m + \Delta u_m, \quad (m = -M, \dots, -1, 0, 1, \dots, M) \quad (28)$$

を得る。

u_m の初期値 $u_m^{(0)}$ としては、簡単に求められ、しかも最終値からの距離が近い、平滑化対数振幅スペクトルのケプストラム c_m を用いる。

式(23)より、

$$\frac{\partial g_k}{\partial u_l} = -\frac{1}{2\pi} \int_{-\pi}^{\pi} f(R(\omega)) e^{j(k-l)\omega} d\omega \quad (29)$$

となるので、式(27)は行列式により、次のように書ける。

$$\begin{bmatrix} a_0 & a_1 & \dots & a_{2M} \\ a_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_1 \\ a_{2M} & \dots & \dots & a_0 \end{bmatrix} \begin{bmatrix} \Delta u_{-M} \\ \vdots \\ \Delta u_0 \\ \vdots \\ \Delta u_M \end{bmatrix} = \begin{bmatrix} b_{-M} \\ \vdots \\ b_0 \\ \vdots \\ b_M \end{bmatrix} \quad (30)$$

ただし、

$$a_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(R(\omega)) e^{jk\omega} d\omega \quad (31)$$

$$b_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(R(\omega)) \cdot R(\omega) e^{jk\omega} d\omega = b_{-k} \quad (32)$$

式(30)は、係数行列が対称テプリッツ行列になっているため、効率的な逐次計算により $O(M^2)$ の積和演算回数で解くことができる⁽⁸⁾。

また、式(30)の係数行列について、 $y = (y_0, y_1, \dots, y_{2M})$ を変数とする 2 次形式 Q をつくると、式(31)より

$$Q = \sum_{k=0}^{2M} \sum_{l=0}^{2M} a_{|k-l|} y_k y_l = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(R(\omega)) |y_0 + y_1 e^{j\omega} + \dots + y_{2M} e^{j2M\omega}|^2 d\omega \quad (33)$$

となる。 $f(x) > 0, x \neq 0$ より、

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} f(R(\omega)) d\omega > 0 \quad (34)$$

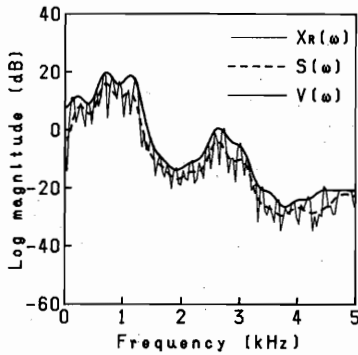
が成り立つので、 $Q > 0 (y \neq 0)$ となり、係数行列は正定値であることが分かる。従って、式(30)の解は必ず存在し、一意に定まる。

更に、式(20)の ε は、式(30)の係数行列 (ε のヘッセ行列) が正定値であるので、Newton-Raphson 法により、最小値に2次収束する⁽⁹⁾。

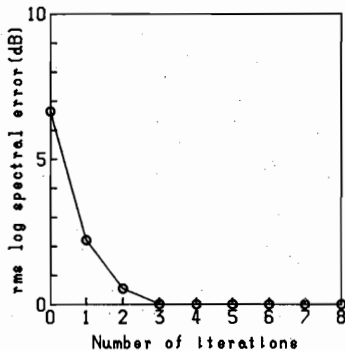
4. 対数スペクトル包絡の抽出例

4.1 収束性

本方法により得られた対数スペクトル包絡の例を図3(a)に、このときの包絡の収束の様子を図3(b)に示す。分析信号は、男性話者による/a/である。標準化周波数は10 kHz、時間窓は25.6 msのブラックマン窓を用い、すべてのフーリエ変換は256点のFFTによ



(a) Smoothed log spectrum $S(\omega)$ and log spectral envelope $V(\omega)$



(b) Log spectral error versus the number of iterations

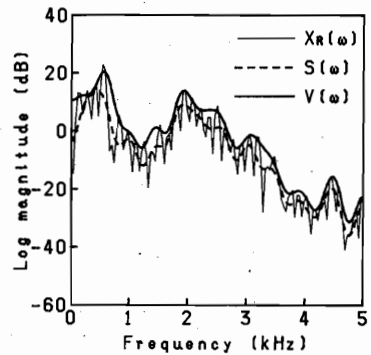
図3 繰り返し計算による収束の様子

Fig. 3 Minimization of the error criterion eq. (21) by the Newton-Raphson method.

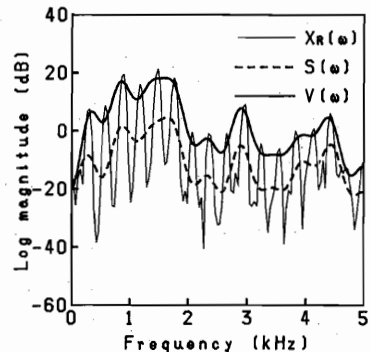
り計算した。また、分析次数は $M=25$ である。図3(a)は8回繰り返し計算したものであり、図3(b)は8回繰り返し計算したものを基準とし、対数振幅スペクトル上での rms 誤差を縦軸にとっている。

σ_1^2, σ_2^2 の比である K を大きな値にするほど、包絡は微細周波数構造のピークに近づくが、収束は K の値が小さい方が早い。ここでは、これらの兼ね合いと、図2のスペクトル残差の分布も考え合わせ、 $K=32$ としている。

Newton-Raphson 法の繰り返しは、ほぼ2, 3回で実用上問題のない程度に収束し、以後の繰り返しにより包絡は、ほとんど変化しないことがわかる。一般的にも、収束の様子は図3とほぼ同じものとなる。従って、通常の繰り返し計算回数は2, 3回程度で十分であり、以下の実験では繰り返し計算回数を2回とする。



(a) Male/e/



(b) Female/a/

図4 本方法による対数スペクトル包絡抽出例

Fig. 4 Examples of log spectral envelopes obtained by the proposed method for voiced sounds.

4.2 スペクトル包絡の抽出例

男性話者による/e/, 女性話者による/a/の分析例を図4に示す。繰り返し計算回数は2回で、その他の分析条件は図3の場合と同じである。ただし、分析次数Mは男性話者によるのがM=25、女性話者によるものがM=20である。いずれも包絡は、微細周波数構造のピーク付近を通っており、良好なものと言える。

無声音に対しては、包絡が微細周波数構造のピークを通るべきだという根拠はないが、包絡のフレームごとの安定性、有声部と無声部での連続性などから、このような仮定は、通常の音声分析において妥当なものと考えられる。

図5に男性話者による無声音/s/の分析例を示す。分析条件は図4(a)の場合と同じである。有声音の場合と同様、上述のような仮定のもとで良好なスペクトル

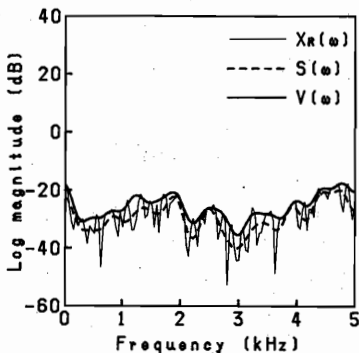


図5 本方法による対数スペクトル抽出例 (無声音/s/)
Fig.5 An example of the log spectral envelope obtained by the proposed method for an unvoiced sound /s/.

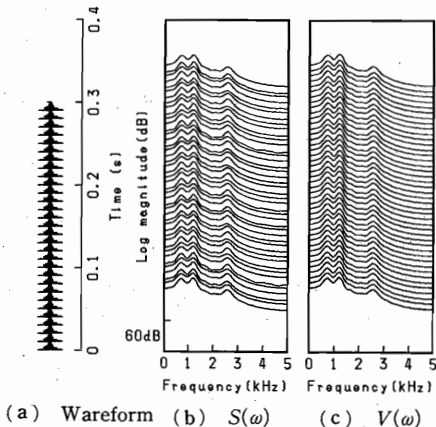


図6 フレーム方向の包絡の安定性
Fig.6 Spectral envelopes for successive analysis frames.

包絡が得られている。

本方法により得られる包絡がフレーム方向にも安定していることを示すため、図6に信号を連続的に分析した例を示す。信号は図2の実験と同じものを用いている。窓の移動周期は6.4 msで、他の分析条件は図4(a)の場合と同じである。

時間的に特性の変化がないデジタルフィルタを、一定レベルのインパルス列で励振しているので、得られる対数スペクトル包絡は、すべてのフレームで等しくなるべきである。しかし、平滑化対数スペクトルS(omega)は、フレームごとにレベルの変動が激しく、包絡の形も一定していない。これが、従来のケプストラム法における合成音声の品質および音声認識率を劣下させる主な原因である。

それに対し、本方法による対数スペクトル包絡V(omega)は、フレームごとのレベル変動がなく、極めて安定しており、包絡の形もすべてのフレームでほぼ同じ形をしている。

5. 他の手法との関係

5.1 改良ケプストラム法⁽⁶⁾

式(20)の最小化問題を最急降下法⁽⁹⁾で解くことを考える。つまり、i番目の近似u_m^(i)に対し、式(20)のepsilonにu_mについて微分し得られる

Delta u_m = - (partial epsilon / partial u_m)
= (1 / (2 * pi)) * integral from -pi to pi of f(R(omega)) * R(omega) * e^j * m * omega * d omega (35)

を計算し、次の近似値

u_m^(i+1) = u_m^(i) + d * Delta u_m,
(m = -M, ..., -1, 0, 1, ..., M) (36)

を得る。ただし、dは、最急降下法における繰り返しの刻み幅である。この場合、式(30)のような線形方程

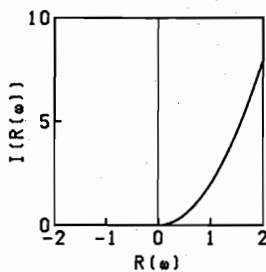


図7 改良ケプストラム法でのスペクトル荷重
Fig.7 The spectral weighting in the improved cepstral method.

式を解く必要がなくなる反面、収束の速度は一般に Newton-Raphson 法と比べ遅くなる。

ここで、 $\sigma_1^2 \gg \sigma_2^2$ より、 $\sigma_1^2=1$, $\sigma_2^2=0$ と近似すると、式(35)、(36)は、対数スペクトル包絡より上に出ている対数振幅スペクトルを平滑化し d 倍したものを前回の近似値に加え込んでいることになる。従って、 $d=(1+\alpha)$ とすれば、改良ケプストラム法となり、式(35)、(36)は式(16)、(17)に等しいことがわかる。

このときの、スペクトル荷重を図7に示す。つまり、改良ケプストラム法は、図7のようなスペクトル荷重による評価を最急降下法により解いていることになる。改良ケプストラム法では、加速係数 α の値を大きくとり過ぎると、包絡が微細周波数構造のピークを越えることがあるが、これは図7のようにスペクトル残差が負の部分で荷重が零になっており、評価式を最小にする解が一意に定まらないためと考えられる。従って、このようなことを避けるためには、式(20) (図1(b)) のようにスペクトル残差の負の部分に小さな重みのかかったスペクトル荷重を用いばよい。

また、収束の速さとの兼ね合いから、最適な加速係数の値は実験的に得られており⁽⁶⁾、この値を用いた場合には、改良ケプストラム法は本方法の近似解法と考えることができ⁽¹⁰⁾、結果的には本方法と非常に近い包絡が得られる。

5.2 重み付き最小2乗法⁽¹¹⁾

式(27)は、重み関数を $f(R(\omega))$ とし、対数振幅スペクトルに重み付き最小2乗法を適用した形になっている。

文献(7)は、重み関数を $r(E(\omega))$ 、ただし、

$$E(\omega) = X_R(\omega) - S(\omega) \quad (37)$$

として、対数振幅スペクトルに重み付き最小2乗法を適用する方法である。従って、重み関数を $f(R(\omega))$ と置き換え、文献(7)の方法を繰り返すことにより、式(18)の誤差 ϵ を最小にする解 u_m を得ることができる。

5.3 ロバスト推定法

ロバスト推定⁽¹²⁾は、ある程度以上の誤差を持つデータは、何らかの原因で誤ったデータであり、誤差評価の対象から外すべきだとする考え方である。式(18)は、包絡より下にあるデータの重みを零に近くすることにより、評価の対象から外しており、負の誤差を持つデータは誤ったデータとしてロバスト推定を適用していると考えられる。

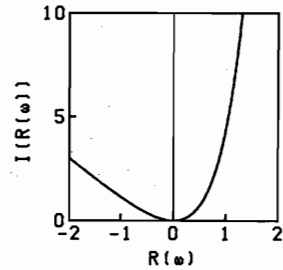


図8 線形予測法でのスペクトル荷重
Fig. 8 The spectral weighting in the linear prediction method.

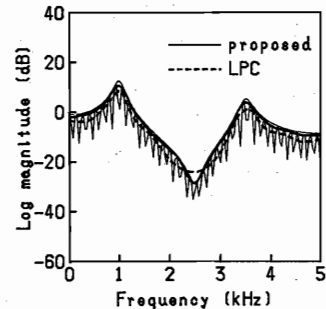


図9 線形予測法による包絡との比較
Fig. 9 Comparison of the linear prediction method and the proposed method.

5.4 線形予測法

線形予測法では、全極モデルのもとで、

$$\epsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\exp(2R(\omega)) - 2R(\omega) - 1) d\omega \quad (38)$$

つまり、図8のような非一様なスペクトル荷重を用いていることになる⁽¹³⁾。全極モデルを仮定した上で、このような荷重を用いているため、包絡はホルマントピークによくマッチするが、零によるスペクトルの谷の部分にはマッチしにくい。

これに対し、本方法では、包絡を極零均等に表現することのできるケプストラムで表しているため、零によるスペクトルの谷を見落とすことはない。

この様子を図9に示す。線形予測法の分析次数は15次で、他の分析条件は図4(a)と同じである。分析信号は、デジタルフィルタを周期7.5 msのインパルス列で励振したもので、フィルタの中心周波数(Hz)/帯域幅(Hz)は、共振点1000/200, 3500/250, 反共振点2500/200である。

6. む す び

非一様なスペクトル荷重により対数スペクトル包絡

を抽出するケプストラム分析法について述べた。

本方法は、

(1) 微細周波数構造のピーク探索あるいはピッチ抽出などを必要としない。

(2) 繰り返し計算による収束が保証されており、2回程度の繰り返しで収束する。

などの特長を持っており、音声の自動分析に適している。実験により、微細周波数構造のピークを通るという意味で良好な対数スペクトル包絡が得られることが確かめられた。本方法により、得られるケプストラム係数を用いることにより、音声合成の品質、音声認識率が著しく向上すると考えられる。

また、他の手法、特に改良ケプストラム法との関係について検討し、本方法は改良ケプストラム法のスペクトル荷重を解が一意に定まるよう修正し、それを最急降下法に替え Newton-Raphson 法で解いていることを示した。

本方法では、分析次数 M の変化に対し、得られる包絡の変化は緩やかなため、分析次数は大まかに定めればよい。しかし、より正確な包絡を得るためには、分析次数の選択が問題となるので、分析次数の自動決定などが今後の研究課題であろう。

文 献

- (1) A. V. Oppenheim and R. W. Schaffer: "Digital Signal processing", Prentice-Hall, Englewood Cliffs, N. J. (1975).
- (2) D. B. Paul: "The Spectral Envelope estimation Vocoder", IEEE Trans. Acoust. Speech and Signal Process., **ASSP-29**, 4, pp. 786-794 (Aug. 1981).
- (3) 三上, 大場: "スペクトル包絡より得られる群遅延特性を用いた有声音の極・零点の推定", 信学論(A), **J68-A**, 1, pp. 86-91 (昭 60-01).
- (4) W. Verhelst, O. Steenhaut: "A New Model for the Short-time Complex Cepstrum of Voiced Speech", IEEE Trans. Acoust. Speech & Signal Process., **ASSP-34**, 4, pp. 43-51 (Feb. 1986).
- (5) D. W. Griffin, J. S. Lim: "A New Model-Based Speech Analysis/Synthesis System", Proc. of ICASSP-85, pp. 513-516 (March 1985).
- (6) 今井, 阿部: "改良ケプストラム法によるスペクトル包絡の抽出", 信学論(A), **J62-A**, 4, pp. 217-223 (昭 54-04).
- (7) T. F. Quatieri: "Minimum and mixed phase speech analysis-synthesis by adaptive homomorphic deconvolution", IEEE Trans. Acoust. Speech & Signal Process., **ASSP-27**, 4, pp. 328-335 (Aug. 1979).
- (8) S. Zohar: "The solution of a Toeplitz set of equations", J. Assoc. Comput. Mach., **21**, 2, pp. 272-276 (April 1974).
- (9) S. L. S. Jacoby 他著, 関根訳: "非線形最適化問題の反復

解法", 培風館 (昭 51).

- (10) 徳田, 小林, 今井: "非一様なスペクトル荷重による対数スペクトル包絡の抽出", 信学技報, **SP86-41** (1986-09).
- (11) 徳田, 小林, 今井: "重み付き最小2乗法による対数スペクトル包絡の抽出", 昭 61 電子通信学会総合全国大会, 1360.
- (12) 中川, 小柳: "最小2乗法による実験データ解析", 東京大学出版会 (昭 57).
- (13) J. D. Markel, A. H. Gray Jr. 著, 鈴木訳: "音声の線形予測" コロナ社 (昭 55).

(昭和 61 年 12 月 23 日受付)



徳田 恵一

昭 59 名工大・工・電子卒。昭 61 東工大大学院修士課程了。現在、同大学院博士後期課程在学中。音声情報処理、デジタル信号処理の研究に従事。



小林 隆夫

昭 52 東工大・工・電気卒。昭 57 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。工博。デジタル信号処理、音声情報処理の研究に従事。日本音響学会、IEEE 各会員。



今井 聖

昭 34 東工大・工・電気卒。昭 39 同大学院博士課程了。同年東工大・精密工学研究所助手。昭 43 同大助教授。昭 54 同大教授。工博。デジタル信号処理、音声の合成および認識の研究に従事。40 年度精機学会論文賞受賞。著書「デジタル信号処理」など。計測制御学会、日本音響学会、IEEE、ASA 各会員。