

解説

音声の特徴抽出と情報圧縮*

板倉文忠** 東倉洋一**

1. ま え が き

音声を対象とする情報処理には、その目的に応じて伝送・蓄積・分析・合成・認識など種々の技術的な側面がある。これらの情報処理を工学的に実現するに、音声波形そのものを処理する代りに、多くの場合音声波形に何らかの前処理をおこない、しかるのちより高次の情報処理を行うのが普通である。このための前処理の部分を音声の特徴抽出と呼び、抽出される特徴要素を特徴パラメータ、あるいは単に特徴と呼ばれる。

音声の特徴抽出の目的は、つぎの2項目に集約されるよう：

- ① 音声波形に含まれる冗長な成分を捨て、有意な情報を選択的に抽出することにより、情報圧縮をおこなう。
- ② 人間の聴覚的な諸属性と、音声波形の物理的な現象ができるだけ整合するような特徴量を選定し、高次の情報処理の実現を容易にする。

音声情報処理の研究においては、②の項目がより重要ではあるが、多くの研究にもかかわらず未だ十分にまとまった成果となっていない。本解説では、主として、①の情報圧縮の面から音声の特徴抽出の代表的な手法について、筆者らの研究、経験を中心に解説してみたい。このため題材の選択には、やや偏りが生じたが、この点の不備に関しては、参考文献^{1)~5)}などにより補って頂きたい。

2. 音声信号モデルと音声の特徴抽出

音声波形は、振幅と位相が時間的にゆるやかに変化する正弦波の和で構成されると考えることができる。音声の知覚において重要な特徴は、主に振幅情報に含

まれており、位相情報の役割は第2義的であるとされている。音声の周波数スペクトルは、時々刻々ゆるやかに変化するが、20~30 ms程度の短時間区間幅では、おおむね一定とみなして取扱うことができる。音声の知覚上重要な特徴は、短時間区間ごとのスペクトル(短時間スペクトル)の時間変化パターンとして表現される。各種の音声特徴抽出は、短時間スペクトルに現われた種々な特性を比較的少数のパラメータによって簡明かつ効率的に表現することをめざしている。

図-1は、男性の発した母音ア/a/の短時間スペクトル分析の一例である。音声の短時間スペクトルは、周波数とともにゆるやかに起伏する成分(スペクトル包絡)と、細かく急速に変化する成分(スペクトル微細構造)の積(dB尺では和)に分解して考えることができる。スペクトル包絡は、発声器官の共振特性のほか、声帯音源波形のスペクトル概形、唇、鼻口からの音波の放射特性などを含めたスペクトルの大域的な概形を表わす部分である。他方、スペクトル微細構造は、大域的にはほぼ平坦なスペクトル分布を示すが、理想的な有声音の場合は均一な高調波成分をもち、また典型的な無声音の場合は白雑音の短時間スペクトルとみなされる。しかしながら現実の音声においては、

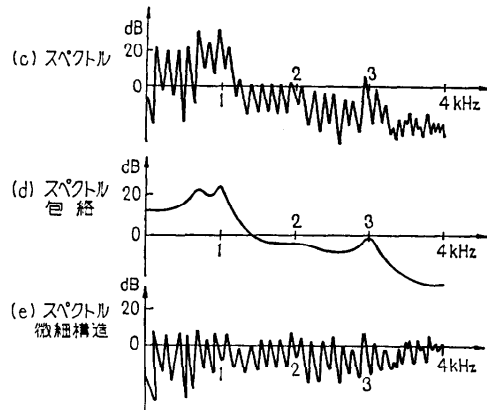


図-1 母音/a/のスペクトル分析

* Feature Extraction of Speech Signal and Its Application to Data Compression by Fumitada ITAKURA and Yoh'ichi TOHKURA (Musashino Electrical Communication Laboratory, N. T. T.).

** 日本電信電話公社武蔵野電気通信研究所

表-1 スペクトル包絡の分析法と特長

分類	分析法	パラメータ	特徴
ノンパラメトリック分析 (NPA)	i) 短時間自己相関分析	$\phi(m)$	スペクトル包絡と微細構造がたまたみこみの形で混在する。計算アルゴリズムが単純でハードウェア化が容易。
	ii) 短時間スペクトル分析	$S(\omega)$	スペクトル包絡と微細構造が積の形で混在。FFT によれば高速処理可能。
	iii) ケプストラム分析	$C(\tau)$	スペクトル包絡と微細構造がケプレンシ領域において近似的に分離。2回のFFTと1回の対数変換必要。
	iv) 帯域フィルタバンク分析	フィルタ出力のRMS値	スペクトル包絡の概形が求まる。アナログ実時間処理に適す。
	v) 零交叉数分析	零交叉数	iv) との組合せによりホルマント周波数も近似的に求まる。ハードウェアが簡単。
パラメトリック分析 (PA)	i) 合成による分析 (Analysis-by-Synthesis)	ホルマント、バンド幅など	モデルの精密化が可能。ホルマントの分析精度が高い。複雑な逐次近似が必要のため、処理時間大。
	ii) 線形予測分析 (LPC)	—	モデルは全極形スペクトルで単純。モデルの次数に等しい遅れの自己相関(共分散)から、逐次近似によらず推定可能。
	a) 最尤スペクトル推定 (自己相関法逆フィルタ法)	α_i	合成フィルタの安定性の保証あり、時間窓が必要。計算量 $O(N^2)$
	b) 共分散法	α_i k_i	合成フィルタの安定性の保証なし。安定化処理が必要。時間窓が不要なため、短時間分析に適す。計算量 $O(N^2)$
	c) PARCO 分析	k_i	正規方程式の求解が格子形フィルタに埋めこまれる。積分特性の変更により、a), b) のいずれとも等価可能。計算量 $O(N^2)$

声帯音源信号の乱れや共振特性の変動のため極めて複雑なスペクトル微細構造が観測されることが多い。このため、音声の特徴抽出にあたっては、これらの複雑な変動に耐え得る手法を確立することが必要である。こうした状況を反映して、最近の音声分析の分野では、既存の統計的信号解析の手法の応用、新しい信号処理方式の提案が行われている。

3. スペクトル包絡パラメータの抽出

多くの統計的信号解析の場合と同様、音声のスペクトル包絡パラメータの抽出法も、ノンパラメトリック分析法(NPA)とパラメトリック分析(PA)に大別して考えられる。NPAは、対象とする音声信号に関してあらかじめモデルを設定せずに行う分析で、自己相関分析、フーリエ分析、帯域フィルタバンクによる分析、ケプストラム分析などがあげられる。NPAは、より広い適用領域をもつ分析法である反面、それにより期待できる情報圧縮の程度もあまり高くないことが多い。

分析対象を音声に限れば、その発生過程がかなり明確な物理系としてモデル化できること、および人間の聴覚がスペクトル上のピーク部(ホルマント)に敏感であることなど、音声に特有の性質に着目してより強い拘束条件をつけたモデルを設定することができる。一般にモデルを規定するパラメータは比較的少数であ

り、これらを最適に推定することにより高精度、高効率なスペクトル情報を抽出することができる。これがPAの基本的な考え方である。歴史的には Analysis-by-Synthesis (A-b-S) がPAの代表例であるが、約10年前に登場した線形モデルにもとづく最尤スペクトル推定法、これと近い関係にある線形予測分析(Linear Predictive Coding, LPC)部分相関分析(Partial Correlation, PARCOR)などが、今日のPAの代表例とすることができよう。表-1は以下にのべるスペクトル包絡の分析法とその特徴をまとめたものである。

3.1 ノンパラメトリック分析

3.1.1 短時間自己相関関数

W Hz に帯域制限した音声波形をナイキスト間隔 $T=1/(2W)$ s で標準化することによって得られる時系列を $x(n)$ ($n=$ 整数) で表わす。短区間の波形の切出しに使う時間窓を $\{w(n)\}_{n=1}^N$ とするとき短時間自己相関関数 $\phi(m)$ は次式で定義される。

$$\phi(m) = \frac{1}{N} \sum_{n=1}^{N-|m|} x(n)w(n)x(n+|m|) \times w(n+|m|), \quad |m|=0, 1, \dots, N-1 \quad (1)$$

$\phi(m)$ を計算する場合、時間窓の形、長さ N および計算精度の選択が問題になる。音声分析においては、時間窓としては⁶⁾、次式で定義されるハミング窓 $w_H(n)$

が好んで用いられる。

$$w_H(n) = 0.54 - 0.46 \cos(n\pi/N) \quad (2)$$

ハミング窓は、周波数領域における分解能が比較的高く、かつサイドロープの減衰が 43 dB 以上あるため、スペクトルの洩れが少ないという利点をもつ。これに対して、矩形窓 $w_R(n) = 1, 1 \leq n \leq N$ は、周波数分解能が一番高いけれども、サイドロープの減衰は 13 dB 程度にすぎないためスペクトル成分の強度範囲の広い音声の分析には適合しない。データ窓の時間長 NT は、30 ms 前後に選ぶのが普通であるが、女声の分析では 20 ms、男声の分析では 40 ms 位に選ぶほうが良い結果をもたらすことがある。入力信号の精度 (ビット数) は、音声信号レベルのダイナミックレンジを考慮に入れると 12~14 ビットの精度がほしいが、短時間区間ごとにブロック正規化をすれば、6~7 ビットにまで精度を減少しても実用上十分な精度の $\phi(m)$ が得られる。

3.1.2 短時間スペクトル

短時間電力スペクトル密度 (略して短時間スペクトル) $S(\omega)$ は $\phi(m)$ とフーリエ変換対をなす。

$$S(\omega) = \frac{1}{2\pi} \sum_{m=-(N-1)}^{N-1} \phi(m) \cos \omega m, \quad (3)$$

$$\phi(m) = \int_{-\pi}^{\pi} S(\omega) \cos \omega m d\omega. \quad (4)$$

ここで、 ω は規準化された角周波数であり、実周波数を f とするとき、 $\omega = 2\pi f T$ と表わされる。実際の $S(\omega)$ の計算では高速フーリエ変換 FFT のアルゴリズムを用いて次式の DFT をもとめる。

$$S(\omega) = \frac{1}{2\pi N} \left| \sum_{n=1}^N w(n)x(n)e^{-j\omega n} \right|^2 \quad (5)$$

ここで、 $w(n)$ はデータ窓であり、 $\phi(m)$ の計算の場合と同様な配慮を払って選択される。

短時間自己相関関数および短時間スペクトルは、ともに原信号の位相情報を消去した表現であるが、この処理による冗長度削減の効果は少ない。しかし、以後にのべるように、広く用いられている特徴パラメータの多くが、 $\phi(m)$ あるいは $S(\omega)$ から誘導されるという意味において、音声の基本的な分析法といえることができる。

3.1.3 ケプストラム分析

ケプストラム (Cepstrum) $C(\tau)$ は、短時間スペクトル $S(\omega)$ の対数の逆フーリエ変換として定義され⁷⁾、 τ は、時間次元でありケプレンシ (quefrency) と呼ばれる。2. で述べたように、 $S(\omega)$ はスペクトル微細構

造とスペクトル包絡の積に分解されるので、 $C(\tau)$ の 0 から 2~4 ms の低ケプレンシ部はスペクトル包絡に対応し、それ以上の τ の部分は主としてスペクトル微細構造に対応する。この性質を利用すると前者からホルマント周波数などのスペクトル包絡情報⁸⁾を、後者からピッチ周期などのスペクトル微細構造情報を抽出することができる⁷⁾。なおケプストラム分析は、2 回のフーリエ変換と 1 回の対数変換を要するため計算量が多いが、専用 FFT プロセッサをもつシステムには適した分析法である。

3.1.4 その他の NPA

スペクトル包絡を正確に抽出するためには、以上のべたような $\phi(m), S(\omega), C(\tau)$ などを計算する必要があるが、帯域フィルタバンクと時間領域の処理の組合せによる比較的少ない計算量の分析法がある⁹⁾。例えば、音声帯域 0.2~6.4 kHz を 5 個のオクターブバンドフィルタに通し、各フィルタの出力の平均振幅 (あるいは RMS 値) と波形の零交叉数を 10 ms ごとに測定する。有声音のばあいは、エネルギーがある閾値を越えたとき、その帯域に 1 つのホルマントを含むとみなす。各帯域内では零交叉数よりホルマント周波数が近似的に求まる。波形ピーク、零交叉数はピッチ抽出や、有声/無声の判定にも利用しやすいパラメータである¹⁰⁾。

3.2 パラメトリック分析

3.2.1 合成による分析, A-b-S

音声生成の音響学的理論によれば、音声信号のスペクトルは音源スペクトル、声道伝送関数および唇、鼻口からの放射特性の和で表わされる。音源と放射特性のスペクトルは比較的声道の形に無関係であるのに対して、声道伝送関数は声道の形と音源位置によってきまり、鼻音化していない母音あるいは母音的な音ではスペクトル極によって、それ以外の音では極と零によって表現できる。

A-b-S は、信号の生成モデルが知られていると仮定して、観測値とモデルにより生成された値の誤差が最小となるようなモデルのパラメータを求め、その値をもってモデルの特性値とする分析法である¹¹⁾。したがって、A-b-S においては、仮定する生成モデル、誤差の評価尺度および誤差の最小化のアルゴリズムの選択が必要である。藤崎らは、母音の線スペクトルの包絡を観測値とし、5 対の極、音源・放射特性ならびに高次の極の補正を 8 個のパラメータで表現し、対数電力スペクトルの平均 2 乗誤差を誤差評価尺度として

用いる方法を提案した¹²⁾。この方法は、以下にのべる線形予測分析に比べて、ホルマントの抽出精度は高いが、誤差の最少化を逐次近似で行うため計算量が格段に多いという欠点があり、主として音声研究における精密測定に用いられる。

3.2.2 線形予測符号化, LPC 分析

線形予測 (Linear predictive coding, 略して LPC) 分析は、音声波形の現時点の標本値 $x(n)$ が過去の p 個の標本値の線形結合で近似される点に着目した分析法である。すなわち、 $x(n)$ は、つぎの差分方程式にしたがって生成されると仮定する。

$$x(n) + \alpha_1 x(n-1) + \dots + \alpha_p x(n-p) = \sigma \varepsilon(n) \quad (6)$$

ただし、 $\sigma \varepsilon(n)$ はモデルの駆動音源入力信号であり、 σ はその RMS 値、 $\varepsilon(n)$ は単位 RMS 値の、平坦なスペクトルをもつ (いいかえれば、近傍標本間で無相関な) 信号であると仮定する。このとき、入力 $\varepsilon(n)$ から出力 $x(n)$ への伝送関数は、

$$H(z) = \frac{\sigma}{1 + \sum_{i=1}^p \alpha_i z^{-i}} \quad (7)$$

で与えられる。LPC 分析は、短時間区間の音声波形から、 $\{\alpha_i\}_{i=1}^p$ および σ 、あるいはそれと等価なパラメータの推定を目的とする。

音声の分野では、1968年音声スペクトル密度の最適識別法に関連して、最尤スペクトル推定法が提案されて以来¹³⁾、今日、一般に LPC 分析と呼ばれる種々の方法が提案されるとともに、理論的な相互のつながりも整理されつつある¹⁴⁾。この節では、これらの代表的なものについてそれぞれの考え方、手法を解説する。

3.2.2.1 最尤スペクトル推定法, 自己相関法

式(6)の駆動音源信号 $\varepsilon(n)$ を理想化して考えると、有声音の場合はインパルス列、無声音の場合は白雑音とみなすことができる。しかしながら、2. に述べたように、現実には有声音の場合でも、無視できない程度のピッチ周期、音源パルス形状のゆらぎが含まれている。最尤スペクトル推定法では、これらの予測困難な変動要因をモデル化して、 $\varepsilon(n)$ は短時間区間で定常的白色ガウス雑音であると仮定する。このように信号に変動要因が含まれることを前提としたうえで、与えられた観測値 (短区間の音声波形) から音声の特徴パラメータ $\sigma, \alpha_1, \dots, \alpha_p$ の最適推定を行おうとするのがこの最尤法である。パラメータ $\Theta = (\sigma^2, \alpha_1, \alpha_2, \dots, \alpha_p)$ のもとの観測値 $X = (x(1), x(2), \dots, x(N))$ の対数尤度は次式で与えられる:

$$\begin{aligned} L(X/\Theta) &= -\frac{N}{2} \left[\log 2\pi\sigma^2 + \frac{1}{\sigma^2} \sum_{i=0}^p \sum_{j=0}^p \alpha_i \phi(i-j) \alpha_j \right] \\ &= -\frac{N}{2} \left[2 \log 2\pi + \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \log T(\omega) \right. \right. \\ &\quad \left. \left. + \frac{S(\omega)}{T(\omega)} \right\} d\omega \right] \end{aligned} \quad (8)$$

ただし、 $\phi(m)$ 、 $S(\omega)$ は、それぞれ短時間自己相関関数およびスペクトラム $T(\omega)$ は、式(7)の $H(z)$ に対応する電力スペクトル密度であり、

$$T(\omega) = \frac{1}{2\pi} |H(e^{-j\omega})|^2, \quad -\pi \leq \omega \leq \pi. \quad (9)$$

Θ の最尤推定値は、 $L(X/\Theta)$ を最大とする Θ として決定される。まず、 $L(X/\Theta)$ の σ^2 について微係数を 0 とおくことにより、

$$\partial^2 = \sum_{i,j=0}^p \alpha_i \phi(i-j) \alpha_j, \quad \alpha_0 = 1 \quad (10)$$

$$\max_{\sigma^2} L(X/\Theta) = -\frac{N}{2} \log(2\pi e \partial^2) \quad (11)$$

さらに、 $\max_{\sigma^2} L(X/\Theta)$ の α_i についての微係数を 0 とすれば、 α_i に関する連立一次方程式 (正規方程式) をうる。

$$\begin{pmatrix} \phi(0) & \phi(1) & \dots & \phi(p-1) \\ \phi(1) & \phi(0) & & \\ \vdots & & \ddots & \\ \phi(p-1) & \dots & \phi(1) & \phi(0) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} = - \begin{pmatrix} \phi(1) \\ \phi(2) \\ \vdots \\ \phi(p) \end{pmatrix} \quad (12)$$

この連立一次方程式の解 α_i を係数とする伝送関数式 (7) は、自己相関行列 $(\phi(i-j))_{i,j=1,p+1}$ の主座行列式が正であれば、常に安定であることが証明される。この条件は、短時間スペクトル $S(\omega)$ がすべての ω について正である事と同等であり、式(1)の $\phi(m)$ を用いればこの条件は満される。

またスペクトル包絡 $T(\omega)$ は、式(7)、(9)に σ, α_i を代入することによりもとまる。また、スペクトル包絡の極は、代数方程式

$$z^p + \alpha_1 z^{p-1} + \dots + \alpha_p = 0 \quad (13)$$

の解としてもとり、対応するホルマント周波数とその帯域幅が決定される。

以上にのべた最尤スペクトル推定法は、入力信号の短時間自己相関関数 $\phi(m)$ にもとづいてパラメータを決定しているので、自己相関法と呼ばれることもある。

なお、 ∂^2 を周波数領域で表わすと

$$\partial^2 = \int_{-\pi}^{\pi} |1 + \alpha_1 e^{-j\omega} + \dots + \alpha_p e^{-j\omega p}|^2 S(\omega) d\omega \quad (14)$$

となり、最尤法は、入力信号を調整可能な p 次の逆フィルタに通しその出力電力 σ^2 を最小にするように係数を調整することと同等である。このため、逆フィルタ法¹⁴⁾と呼ばれることもある。

さて、式(8)の対数尤度を、 $T(\omega) > 0$ という条件のみのもとで $T(\omega)$ に関して最大化すると、 $T(\omega) = S(\omega)$ のとき、 $L(X/\theta)$ は最大値 L_{\max} をとり、 $L_{\max} - L(X/\theta)$ は、次式の $E(T(\omega)/S(\omega))$ に比例することが示される。

$$E_1(T(\omega)/S(\omega)) = \int_{-\pi}^{\pi} 2 \left[\log \frac{T(\omega)}{S(\omega)} + \frac{S(\omega)}{T(\omega)} - 1 \right] d\omega \quad (15)$$

当然 E_1 は、 $T(\omega) \triangleq S(\omega)$ のときに限り零となり、それ以外のとき正となる。いいかえれば E_1 は、入力の短時間スペクトルを、 p 次の全極形スペクトルモデルでおきかえたときの両者の誤差評価尺度である。 $T(\omega)$ と $S(\omega)$ の対数スペクトルの差を $D(\omega)$ とすると、式(15)の非積分関数は

$$2[D(\omega) + \exp(-D(\omega)) - 1] \quad (16)$$

となり、 $|D(\omega)| \ll 1$ の場合は、近似的に $D(\omega)^2$ となり、A-b-S において用いられる誤差評価尺度と近似的に等しい。いいかえれば、最尤スペクトル推定法は、音声の生成モデルを全極形スペクトルモデルで表わしたときの A-b-S と言うことができる。

スペクトル包絡を全極形に限定し、かつ音源のスペクトルを平坦とみなすことにより、A-b-S のように逐次近似によらずパラメータを直接的かつ比較的少ない計算量で求めることができる。

3.2.2.2 共分散法¹⁵⁾

自己相関法が、スペクトル領域での A-b-S として説明できるのに対して、共分散法は波形領域での A-b-S とみなすことができる。 $x(n)$ を $x(n-1), \dots, x(n-p)$ の線形結合 $\hat{x}(n)$

$$\hat{x}(n) = - \sum_{i=1}^p \alpha_i x(n-i) \quad (17)$$

で近似するとき、 $x(n)$ と $\hat{x}(n)$ の誤差評価尺度を

$$E_2 = \sum_{n=1}^N (x(n) - \hat{x}(n))^2 \\ = \sum_{n=1}^N (x(n) + \dots + \alpha_p x(n-p))^2 \quad (18)$$

に選んで、 E_2 を最小化のため α_i について微分して 0 とおけば、つぎの正規方程式をうる。

$$\sum_{j=1}^p \alpha_j \phi(i, j) = -\phi(i, 0) \quad (i=1, \dots, p) \quad (19)$$

ただし、 $\phi(i, j)$ は、 $x(n)$ の共分散関数であり、

$$\phi(i, j) = \frac{1}{N} \sum_{n=1}^N x(n-i)x(n-j) = \phi(j, i). \quad (20)$$

このように、共分散法は、 $x(n)$ が式(17)の $\hat{x}(n)$ で近似できるという線形予測モデルを基盤においた波形領域における一種の A-b-S と言うことができる。

3.2.2.3 部分自己相関法, PARCOR 分析

最尤スペクトル推定法および共分散法が、振幅 σ 、予測係数 α_i を推定するのに対し、部分自己相関 (Partial auto-correlation, 略して PARCOR) 法は、線形予測操作と相関操作を組合せにより予測誤差の相関係数の形でスペクトル包絡情報を抽出する方法である¹⁷⁾。

$\{x(n-i)\}_{i=1}^{m-1}$ から $x(n)$ および $x(n-m)$ を最小 2 乗近似したときの誤差を、それぞれ次式で表わす。

$$f^{(m-1)}(n) = \sum_{i=0}^{m-1} \alpha_i^{(m-1)} x(n-i) = A^{(m-1)}(D)x(n) \quad (21)$$

$$g^{(m-1)}(n) = \sum_{i=1}^m \beta_i^{(m-1)} x(n-i) = B^{(m-1)}(D)x(n) \quad (22)$$

ただし、 $A^{(m-1)}(D), B^{(m-1)}(D)$ は、それぞれ $m-1$ 次の前向および後向の予測誤差演算子であり、 α, β を係数とする単位遅延 D に関する多項式である。このとき、時間差 m における PARCOR 係数 k_m は $f^{(m-1)}(n), g^{(m-1)}(n)$ 間の相関係数として定義される。 $m=1$ のときは $f^{(0)}(n) = x(n), g^{(0)}(n) = x(n-1)$ ゆえ、 k_1 は通常の自己相関係数に等しい。 $m=2$ のときは、 $f^{(1)}(n) = x(n) - k_1 x(n-1), g^{(1)}(n) = x(n-2) - k_1 x(n-1)$ であり、 $f^{(1)}(n)$ は $x(n)$ から $x(n-1)$ による予測値がさし引かれたもの、 $g^{(1)}(n)$ は $x(n-2)$ から $x(n-1)$ による予測値がさし引かれたものであり、 k_2 は両者の間の相関係数として求まる。このように、 $\{k_m\}_{m=1}^p$ は、予測による相関除去と、前段で除去つくされない相関の検出というつの操作を繰返すことにより逐次抽出される。この操作を 10 回前後繰返すことにより、 $f^{(p)}(n), g^{(p)}(n)$ はほぼ平坦なスペクトルをもつ信号に変換され、そこには主に駆動音源の振動とスペクトル微細構造(周期性の有無、ピッチ周期)に関する情報のみが残されることになる。この性質に着目して $f^{(p)}(n)$ の自己相関関数から音源情報を抽出することができる。

PARCOR 分析により得られた特徴パラメータからの音声の復元合成は、分析と逆の過程により実現される。すなわち、有声音の場合はピッチ周期の等しいインパ

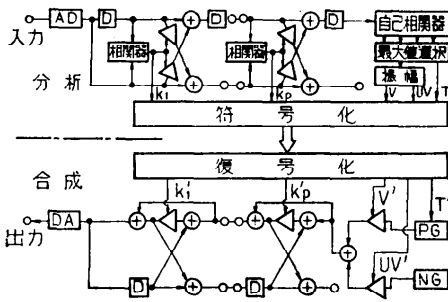


図2 PARCOR 形音声分析合成系

ルス列，無声音の場合は白雑音により駆動音源信号を発生させ，分析における予測演算子 $A_p(D)$ の逆特性をもつデジタルフィルタを駆動することにより，原音声に近いスペクトルの音声合成される(図-2)。

3.2.2.4 自己相関法，共分散法・PARCOR 分析の比較

PARCOR 分析は，予測誤差 $f^{(m-1)}(n)$ と $g^{(m-1)}(n)$ の相関係数を求める場合の時間窓の形あるいは積分特性の選び方によって，自己相関法，共分散法，あるいは両者の中間的なものに等価な分析法を考案することができる¹⁸⁾。そこで本節では，LPC 分析の代表例である自己相関法と共分散法について両者の比較を行う。

予測係数を計算するための両方法の正規方程式式(12)，(19)を比較すると，前者では単に係数行列の要素が $|i-j|$ だけにより，後者では単に係数行列が対称なだけである。これらの性質により，自己相関法では，Durbin の方法¹⁹⁾によって a_i, k_i, σ が同時に求まり，それに要する乗算・除算の回数は，それぞれ

p^2, p 回である。また，共分散法では，Cholesky の方法¹⁵⁾による乗算・除算，開平の回数は，それぞれ， $(p^3+9p^2+2p)/6, p, p$ 回であり，仮に $p=10$ とすれば，両者の間に約3倍の計算量の差がある。なお， $\phi(m), \phi(i, j)$ の計算は，整数演算に適しており，比較的簡単なハードウェアにより実時間処理が可能であり，両者の差は少ない。

計算量をのぞいた自己相関法と共分散法の優劣は，その適用分野に依存するので一概に論ずることは困難である。ここでは，分析結果の一例として図-3に2方法によるスペクトル包絡の抽出例を示す。30ms のハミング窓を用いた自己相関法によれば，同じ長さのデータによる共分散法による場合よりもより滑らかにスペクトル包絡が抽出されていることが知られる。

4. 音源パラメータの抽出

声帯音源信号のもつスペクトル微細構造を表すパラメータを音源パラメータと呼ぶ。音源信号のパラメトリック分析では，音声の発声モデルに基づき，音源を2つの型に分類する。その一つは有声音源で，声帯の振動による呼気の断続によって作られる脈動流(パルス)であり，他の一つは無声音源で，声道が狭くなった所で生ずる空気の流れ(ノイズ)である。音源分析では，音声の有声音 V (voiced) か無声音 UV (unvoiced) かを決定し，有声音の場合には脈動の周期であるピッチ周期を求める。すなわち，有聲/無聲 (V/UV) とピッチ周期 (T) の二つが音源パラメータである。この音源分析をピッチ抽出法と呼び，V/UV の決定は，このピッチ抽出法に付随したものと扱われるのが

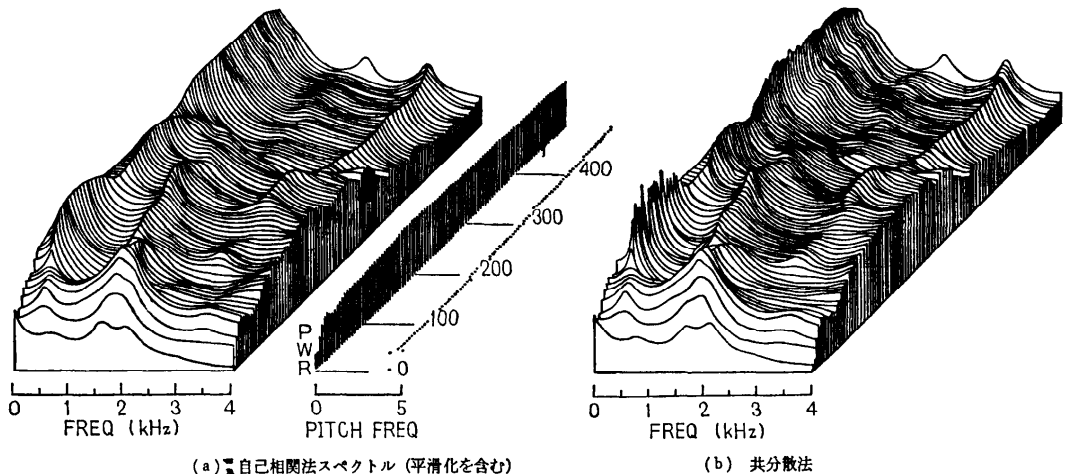


図-3 音声の線形予測分析における自己相関法と共分散法の比較

普通である。

4.1 ピッチ抽出

ピッチ抽出は、音声分析の研究が始まって以来の研究課題であり、今なお、決定的な方法は報告されていない。ピッチ抽出の数多い研究のなかの代表的なものを表-2に示す。ピッチ抽出法は、(I)波形処理、(II)相関処理、(III)スペクトル処理、に大別される。波形処理を使った抽出法は、図-4(a)の音声波形にみられるピッチパルスに該当する波形の周期的ピークを検出する方法である。このピーク検出手法として、(a)並列処理法¹⁰⁾(Parallel Processing)、(b)データ減少法¹⁹⁾(Data Reduction)、(c)零交叉計数法²⁰⁾(Zero Crossing Count)などがある。

相関処理を使ったピッチ抽出は、音声のデジタル信号処理において最も広く用いられている。これは、相関処理が波形の位相むずみに強く、また、ハード化においても比較的簡単な構造(高速度の処理が要求される)で実現できるためである。相関処理は、音声波形の自己相関関数を用いるものと^{21)~23)}、線形予測分析における残差信号の自己相関関数を用いる変形相関法¹³⁾に大別できる。さらに、相関処理の方法論の問題として、自己相関関数の代りに平均振幅差関数^{24),26)}

表-2 ピッチ抽出法と特長

分類	ピッチ抽出法	特長
I. 波形処理	i) 並列処理法 (Parallel Processing)	6種類の簡単な波形ピーク検出器で抽出されたピッチ周期の多数決。
	ii) データ減少法 (Data Reduction)	種々の論理操作により、波形データからピッチパルス候補以外のデータを捨てていく。
	iii) 零交叉計数法 (Zero Crossing Count)	波形の零交叉数に関する繰返しパターンに着目。
II. 相関処理	iv) 自己相関法	音声波形の自己相関関数。センダークリップによりスペクトルを平坦化、ピーククリップにより演算の簡略化を計る。
	v. a) 変形相関法	LPC分析の残差信号の自己相関関数、残差信号のLPFと極性化により演算の簡略化が可能。
	v. b) SIFT アルゴリズム	音声波形のダウンサンプリング後、LPC分析を行い、逆フィルタによってスペクトルを平坦化、相関関数の補間により時間精度を回復。
	vi) AMDF	平均振幅差関数 (AMDF) によって周期性を検出。残差信号の AMDF による抽出も可能。
III. スペクトル処理	vii) ケプストラム	パワースペクトルの対数のフーリエ変換によりスペクトルの包絡と微細構造を分離。
	viii) ピリオドヒストグラム	スペクトル上で、基本周波数の高調波成分のヒストグラムを求め、高調波の公約数によってピッチを決定。

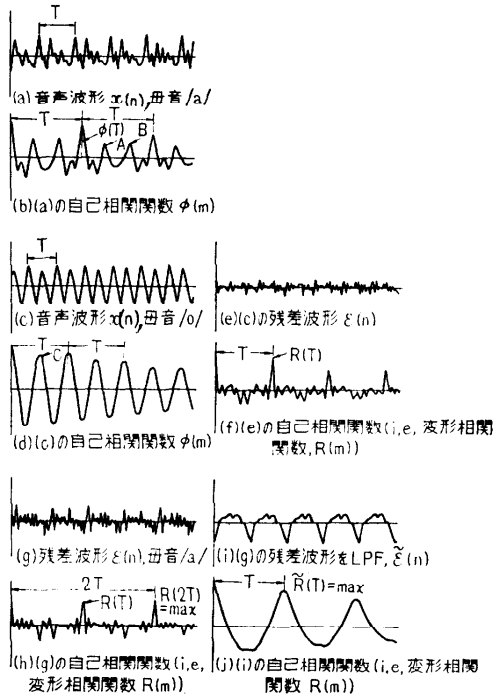


図-4 音声波形 $x(n)$ 、残差波形 $\epsilon(n)$ および自己相関関数 $\phi(m)$ 、変形相関関数 $R(m)$

(AMDF, Average Magnitude Difference Function) を用いる方法や、自己相関演算を簡略化した極性相関法²⁶⁾などに区別される。また、変形相関法の前処理として、音声サンプルのダウンサンプリングを行い、計算量の節約を狙ったものに、SIFT アルゴリズム²⁷⁾がある。

スペクトル処理によるピッチ抽出の代表例が、ケプストラム分析⁷⁾によるピッチ抽出である。ケプストラム分析により、スペクトル包絡構造とスペクトル微細構造が分離されるから、有声音の場合のケプストラムは、ピッチ周期に相当するケプレンシ、 T において強いピークをもち、ピッチ周期 T が抽出される。ピリオドヒストグラム²⁸⁾も、音声スペクトルに着目したピッチ抽出法である。この方法は、ピッチ周波数の高調波成分のヒストグラムを求め、このヒストグラムに基づいて、高調波周波数の公約数として基本周波数を求める方法である。以上、ピッチ抽出法を概観した。このなかで相関処理を使ったピッチ抽出法に関して、(i) 音声波形の自己相関関数と残差波形の自己相関関数、(ii) 自己相関関数と AMDF および極性相関、に絞って、4.2、4.3 に詳しく説明する。

4.2 自己相関関数と変形相関関数

図-4 に、音声波形 $x(n)$ とその自己相関関数 $\phi(m)$ を示す。 $\phi(m)$ は、ピッチ周期 T に一致する点において、顕著なピーク値 $\phi(T)$ を持つから、このピーク値を検出し、ピッチ抽出を行えばよい。しかし、 $\phi(m)$ には、ホルマントの影響によって $\phi(T)$ 以外のピーク (図-4(b)の A, B) が存在する。図-4(c) 母音 /O/ の例では、このホルマントの影響によって、 $\phi(T)$ と同等の値を持つピーク C が表われている。このようなホルマントの影響は、ピッチ抽出誤差の原因となるから、このホルマントの影響を除去し、スペクトルを平坦化する前処理が必要となる。このスペクトル平坦化には、波形のセンタ、クリッピングやフィルタバンク^{21), 22)}が使われている。

一方、残差信号 $\varepsilon(n)$ は、適当な予測次数の線形予測分析によって、すでに、スペクトル包絡がほぼ平坦化されている。この $\varepsilon(n)$ の自己相関関数 $R(m)$ を変形相関関数と呼ぶ。すなわち、 $R(m)$ においては、線形予測分析によってホルマントの影響が除去されており、周期性を表わすピークだけが明りょうになる (図-4(f))。

変形相関関数 $R(m)$ を求める方法は2通りある。一方は $\varepsilon(n)$ の自己相関関数を求める方法であり、この方法は、分析過程で $\varepsilon(n)$ が求められている格子形の PARCOR 分析と組合せると有効である。他方は、波形の自己相関関数 $\phi(m)$ にスペクトル包絡除去操作を行い、 $R(m)$ を求める方法である。 $R(m)$ の定義により、

$$R(m) = \frac{E \{ \varepsilon^{(p)}(n) \varepsilon^{(p)}(n+m) \}}{E \{ \varepsilon^{(p)}(n)^2 \}} \\ = \frac{1}{u_p} E \left\{ \sum_{i=0}^p \alpha_i^{(p)} x(n-i) \sum_{j=0}^p \alpha_j^{(p)} x(n+m-j) \right\} \quad (23)$$

$$= \sum_{i=-p}^p A_i^{(p)} \phi(i+m) \Big/ \sum_{i=0}^p \alpha_i^{(p)} \phi(i) \quad (24)$$

ただし、

$$A_i^{(p)} = \sum_{j=0}^{p-|i|} \alpha_j^{(p)} \alpha_{j+|i|}^{(p)}, \quad i=0, \pm 1, \dots, \pm p. \quad (25)$$

$$u_p = E \{ (\varepsilon^{(p)}(n))^2 \} = \sum_{n=0}^p \alpha_n^{(p)} \phi(i) \quad (26)$$

となる。すなわち、 $R(m)$ を求めるには、波形の自己相関関数 $\phi(m)$ に、重みを $\{A_i^{(p)}\}_{i=-p}^p$ とする荷重移動平均操作を行えばよい。この方法は、PARCOR

分析を自己相関法で行う方法と組合せる場合に有効である。図-4(e), (f)は、(c)の $x(n)$, (d)の $\phi(m)$ に対応した $\varepsilon(n)$, $R(n)$ を比較したものである。 $R(m)$ では、 $\phi(m)$ において観察されたホルマントの影響が取除かれ、 $R(T)$ だけが明りょうなピークとして抽出される。

このような相関処理においては、相関演算の前処理として、信号を低域濾波 (f_c = 約 900 Hz) し、高次ホルマントの影響を取り除いておくことが効果的である。この低域濾波は、変形相関の時間分解能の不足を補う上でも有効となる²⁶⁾。例えば、図-4(h)の $R(m)$ において、ピッチ周期として検出されるべき第一ピーク $R(T)$ の先端が欠除し、 $R(2T)$ が最大値を示す。この結果、周期 $2T$ をピッチ周期とする誤抽出を招く。これに対し、低減濾波された残差信号 $\varepsilon(n)$ (図-4(i)) から求めた変形相関関数 $\hat{R}(m)$ (図-4(j)) では、 $R(T)$ が最大値となり、正しいピッチ抽出が行われる。

4.3 AMDF と極性相関関数

自己相関関数を求める演算は、一つの乗算器と一つの加算器による構成が可能であるが、処理の高速化が要求される。このため、演算処理を更に簡略化することを目的とした方法が、AMDF^{24), 29)} と極性相関関数である。

AMDF, $D(m)$ は、

$$D(m) = \frac{1}{N} \sum_{n=1}^N |x(n) - x(n-m)| \quad (27)$$

であり、自己相関演算における乗算を減算に置き換え、演算処理の高速化に応じたものである。したがって、 $D(m)$ は $m=T$ においてディップ値 $D(T)$ を持ち、この $D(T)$ を検出することによってピッチ抽出が行われる。AMDF の詳細は、文献 24) にゆずる。

極性相関関数は、信号 $x(n)$ を極性化 (i.e. $x(n) = \pm 1$ if $x(n) \geq 0$, 複号同順) し、この極性信号の自己相関関数として定義されるから、相関演算に乗算を必要としない。以上、AMDF と極性相関関数は、音声波形に対してと同様、残差信号の処理に利用することもできる^{25), 26)}。

ここに取り上げたピッチ抽出法は、抽出誤りを修正するための適切な論理操作と組合せることにより、おおむね、良好な結果を与えることが報告されている³⁰⁾。今後は、新しいピッチ抽出法の研究と共に、電話音声や、環境雑音を含んだ音声に対し、安定に動作するピッチ抽出法の研究が必要であろう³¹⁾。

4.4 有声/無声 (V/UV) の決定

V/UV の決定は、ピッチ抽出法と結びつけられる場合が多い。例えば、自己相関関数や変形相関関数のピーク値、 $\phi(T)$ 、 $R(T)$ の大ききで V/UV の決定を行う。すなわち、 $\phi(T)$ 、 $R(T)$ が大きい場合には、周期性が大であり、これを有声とし、 $\phi(T)$ 、 $R(T)$ が小さい場合には無声とする。このようなピッチ抽出法と組合せた V/UV の決定は、周期性ではないが有声である部分の V/UV 決定において有効に働かえない。したがって、V/UV の性質と関連の深い他のパラメータを決定の補助パラメータとして用いる³²⁾。この補助パラメータには、i) 音声信号のエネルギー、ii) 零交差の数、iv) 第1次の自己相関係数、vi) 第1次の線形予測係数 (α_1)、v) 残差信号のエネルギー、などがある。最近では、これら5つのパラメータの V/UV に関する統計的性質を使った V/UV の識別法も報告されている³³⁾。

5. 特徴パラメータからの音声の復元

音声の PA 分析で求められた特徴パラメータからの音声の復元は、発声モデルに基づいた形をとる。図-5 に示すように、まず、音源分析で求められた V/UV とピッチ周期にしたがって音源信号が合成される。この音源信号は、PA 分析で抽出されたスペクトル包絡パラメータによって決定される共振特性によってコントロールされ、原音声は復元される。この声道の共振特性を電気回路で実現するため、ターミナルアナログや声道アナログなどの方法が行われた³⁴⁾。

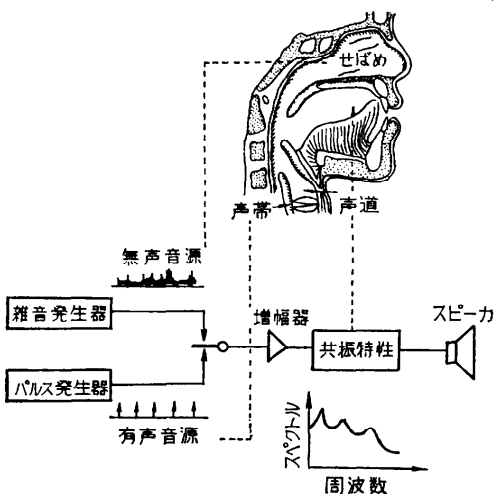


図-5 特徴パラメータからの音声の復元

分析合成系における音声の復元では、それぞれの分析方式で求めたパラメータに適した合成方式が用いられる。例えば、LPC ボコーダの合成部も、伝送パラメータが線形予測子 (α -パラメータ) の場合は、トランスバースフィルタを使った合成¹³⁾、PARCOR 係数 (k -パラメータ) の場合には、格子形のデジタルフィルタを使った合成が適する¹⁷⁾。

図-2 は、PARCOR 分析合成方式における合成器の基本構成の一例である。パルス発生器によって作られる有声音源は、その周期がピッチ周期によりコントロールされる。無声音源は、白雑音発生器であり、音源分析で求められた V/UV パラメータによって、両音源の選択が行われる。このようにして作成された平坦なスペクトルを持つ音源信号に対し、 k -パラメータでコントロールされるデジタルフィルタによって共振特性が与えられる。

以上は、パラメトリックな音源分析でえた音源パラメータ (V/UV とピッチ周期) を使った合成であるが、このような音源パラメータの抽出を行わず、残差信号波形を直接符号化し、伝送する方式^{34)~36)} も報告されている。

6. パラメータの符号化と情報圧縮への応用

音声の PA 分析は、音声の特徴抽出であると共に、情報圧縮の手段ともなる。分析合成系、すなわち、音声波形をスペクトル情報と音源情報にわけて伝送し、合成により波形を再生する方法をボコーダ方式と呼ぶ。各種ボコーダ方式に関しては文献 1) および 3) にまとめられている。分析合成系を利用した音声の情報圧縮の研究は、近年、線形予測分析が提案されて以来、非常に盛んとなった。LPC ボコーダを例にとった場合、各方式の伝送容量は、600~数千 bps と、多種多様である。適用範囲が異なり、方式も微妙な相違を示す各方式を直接比較することは困難であるが、これらの方式のなかで実用上注目されているのが、2,400bps 程度の伝送容量を持つデジタル低ビット伝送である。

一般に、LPC ボコーダの分析部では、30 ms 前後の長さを持つ時間窓を使って、音声波形を切り出し、スペクトル情報を担う LPC パラメータの抽出を行う。この切り出された音声波形をフレームと呼ぶ。音声の情報圧縮を行うためには、許容音声品質において、i) フレーム当りのスペクトル情報および音源情報を減少させること、ii) フレーム更新周期を長くするこ

と、の2点が主体となる。前者の目的のために必要とされるのが LPC パラメータの最適量子化である。後者のフレーム更新周期は、スペクトル情報の時間的な精度に係わるものであり、その長さは、通常の発声速度のもとでは、20~30 ms が限度となる。

スペクトル情報を担うパラメータの最適符号化において重要なことは最適符号化の基準である。すなわち、パラメータの符号化によって生ずるスペクトルのひずみを最小に抑えると同時に、このスペクトルひずみが与える知覚的影響を最小にするような符号化が望まれる。

PARCOR 分析合成方式における k -パラメータの最適符号化に話題を絞る。 k -パラメータの最適符号化に関する研究は数多い³⁷⁾が、ここでは、通話品質を表す AEN³⁸⁾ (明りょう度等価減衰量) とよく対応した評価尺度であるスペクトルひずみを使った最適符号化法を説明する。

6.1 不均一ビット配分

パラメータの量子化によるスペクトルひずみ DS を

$$DS = \frac{1}{2\pi} \int_{-\pi}^{\pi} [S'(\omega) - S(\omega)]^2 d\omega \quad (28)$$

と定義し、全体のスペクトルひずみを、ひずみの時間的な平均で表す。ただし、 $S'(\omega)$ は、パラメータの量子化ひずみを持つ対数スペクトルであり、また、 DS は、通話品質を表す AEN と線形な関係にあることが知られている³⁹⁾。

いま、パラメータ k_i だけを q_i ビットに量子化し、他のパラメータを量子化しない時の DS を $DS_{q_i}^{(i)}$ (q_i) で表せば、

$$DS^{(i)} = DS_{q_1}^{(i)}(q_1) + DS_{q_2}^{(i)}(q_2) + \dots + DS_{q_p}^{(i)}(q_p) \quad (29)$$

となり、ひずみの加算則が成立する。一方、各 k パラメータの生む量子化ひずみは不均等であり、

$$DS_{q_1}^{(i)} > DS_{q_2}^{(i)} > \dots > DS_{q_p}^{(i)}$$

なる関係がある。したがって、パラメータの量子化ビット数 $N (= \sum_{i=1}^p q_i)$ を各パラメータに均一に割当てるよりも、不均一 ($q_1 > q_2 > \dots > q_p$) に割当てた方が、ひずみ DS を小さくすることができる。この DS の最小化を動的計画法を使って行い、その結果求められたビット割当てが、不均一ビット配分である³⁹⁾。

6.2 パラメータの非線形変換

k パラメータによる音声スペクトルの復元精度は、周波数によって不均一である。言い換えれば、 k パラメータのスペクトル感度が異なるため、 k_i に対する

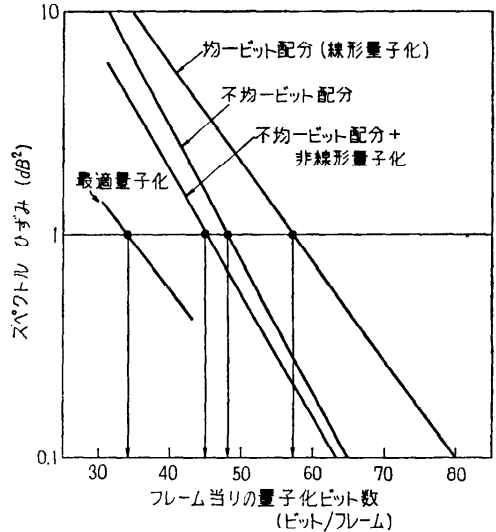


図-6 PARCOR 係数 $\{k_i\}_{i=1}^p$ の量子化とスペクトルひずみ

微小変化 Δk_i の生むスペクトルひずみも一様ではない。このため、 k パラメータを適当に非線形変換することによるスペクトル感度の均一化が効果的となる⁴⁰⁾。この非線形変換には、 $\theta_i = \arcsin k_i$ 、 $q_i = (1/2) \ln((1+k_i)/(1-k_i))$ (対数断面積比) が用いられ、特に低次の PARCOR 係数 k_1, k_2 の量子化特性改善に有効となる。

スペクトルの量子化ひずみを 1 dB 以下に抑えることを基準とした場合、 k パラメータの量子化に必要なフレーム当りのビット数は、不均一ビット配分のときに約 48 ビット/フレーム、これに非線形変換を併せ考慮したときに約 45 ビット/フレームとなる(図-6)、これらの符号化は、均一ビット配分の線形量子化の場合の約 57 ビット/フレームと比較した場合、非常に大きい情報圧縮効果となっている。しかし、低ビット伝送におけるフレーム・レート、 F_s を 50 Hz とすれば、45 ビット/フレームの符号化を使った場合の伝送容量は、 k パラメータの伝送だけで、2,250 ビット/秒となる。音源パラメータの伝送に必要な情報量は、フレーム当り約 12 ビット (600 ビット/秒、 $F_s=50$ Hz) であるから、低ビット伝送における k パラメータの符号化法として、更に能率的なものが要求される。

6.3 パラメータの統計的性質

k パラメータは、それぞれ相異なる統計的性質をもつ⁴⁰⁾。各パラメータが固有の統計的分布を示す場合の量子化においては、すべてのパラメータに対して一様

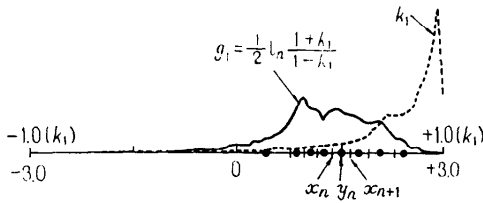


図-7 パラメータの統計的性質を考慮した量子化

な量子化特性を設定するよりも、それぞれのパラメータに適した量子化特性を使った符号化が効果的である。この統計的性質を考慮した符号化の例を、 k_1 の符号化に関して説明する。図-7の実線は、 k_1 を対数断面積比 g_1 に変換し、この g_1 に関して求めた頻度分布である。この g_1 の分布関数を $P_1(x)$ とし、分布を考慮した量子化特性をつぎのように決定する。いま、 $x_n < x < x_{n+1}$ の範囲の x を y_n とする量子化において、量子化の2乗誤差は

$$D_1 = \sum_{n=1}^N \int_{x_n}^{x_{n+1}} (x - y_n)^2 P_1(x) dx \quad (30)$$

となる。ここで、 N は量子化のステップ数である。この量子化誤差を最小にする条件は、 $\partial D_1 / \partial x_j = 0$ ($j = 2, \dots, N$), $\partial D_1 / \partial y_j = 0$ ($j = 1, \dots, N$) となり⁴³⁾、初期値 $x_1 = -3$ として、 y_1 を適切に選べば、 x_n, y_n が逐次決定される。図-7の例は、 k_1 の3ビット量子化において、上記の方法で逐次決定された x_i ($i = 1, \dots, 9$), y_i ($i = 1, \dots, 8$) を示す。

この統計的性質を考慮した符号化を、不均一ビット配分および非線形変換と併せ用いることにより、フレーム当りの量子化ビット数約 34 ビットで、スペクトルひずみ 1 dB を達成することができる(図-6)。したがって、パラメータの分布を考慮することによる情報

圧縮効果は大きく、 $F_r = 50$ Hz において、 k パラメータの伝送に 1,700 ビット/秒、音源パラメータの伝送に必要な情報量を併せても 2,300 ビット/秒の低ビット分析合成系が実現可能となる。

6.4 スペクトル平滑化による量子化特性の改善

k パラメータの量子化によって生ずるスペクトルひずみは、分析合成系への入力音声、すなわち、発声者依存性を持つ⁴²⁾。一般に、ピッチの高い女声においては、このスペクトルひずみが大きい。このスペクトルひずみの発声者依存性は、前項の最適符号化によっても完全に解決されるには到らない。

ピッチの高い女声の分析では、スペクトルの推定に与えるピッチ周期の影響が大きく、ホルマントバンド幅の過少推定により、 k パラメータ、とくに k_1 のスペクトル感度を著しく高めることがある。このような感度の高い部分では、 k_1 の量子化に伴うスペクトルひずみが、他の部分に比べ非常に大きい値をとる。したがって、この問題を解決するためには、著しく感度の高い部分に対する適切な感度低下作用が必要となる。このスペクトル感度低下の手段が、PARCOR 方式におけるスペクトル平滑化⁴³⁾であり、自己相関関数にラグ窓を乗ずることによって実現される。このスペクトル平滑化を最適符号化と併せ用いることによる量子化ひずみの改善効果は大である⁴³⁾。図-8は、(a) k パラメータの線形量子化と(b)最適な量子化とスペクトル平滑化の併用を比較したものであり、音声サンプルや分析の条件は、図-3 のものと同一である。双方とも、PARCOR 係数 $\{k_i\}_{i=1}^1$ フレーム当りの情報量が 40 ビットの場合であり、それぞれのスペクトルひずみは、式(28)の定義にしたがって(a)が約 6 dB²、(b)が約 0.6 dB² となる。分析合成系において

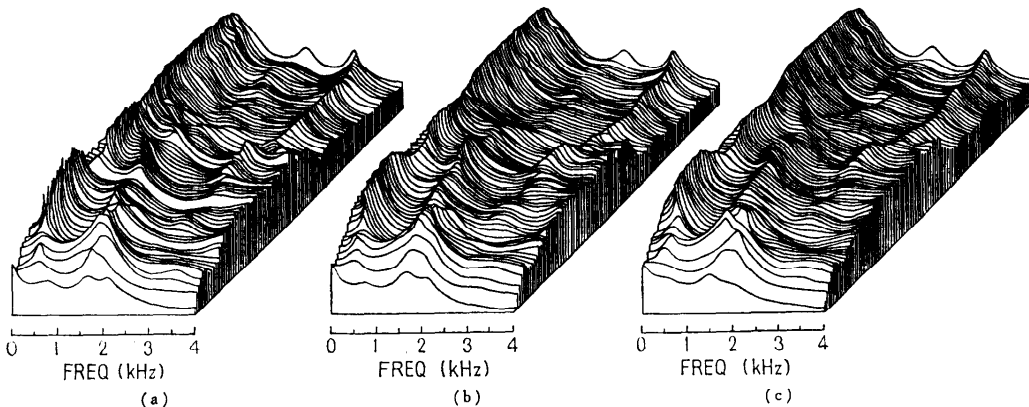


図-8 k パラメータの最適量子化によるスペクトルひずみの低減 (40 ビット/フレームの量子化)

は、合成音声の品質向上の目的で、(b)のスペクトル包絡に対し、時間的な平滑化を行うことが効果的である。(b)において最適量子化された k パラメータを遮断周波数 50 Hz の LPF に通し、時間的に平滑化したものが(c)である。(c)に示されたスペクトルは、種々のひずみの改善が行われた結果、図-3(b)の推定スペクトルの原形に近いものとなっている。

7. む す び

以上、音声の特徴抽出に関して、線形予測分析を中心に各種の手法を紹介した。音声の特徴のうち、スペクトル包絡情報に関しては、かなり組織的な研究が既に行われているが、音源情報(ピッチ, V/UV)に関しては未整理な部分が多く、今後の研究に期待される部分が多い。

音声の特徴抽出処理は、従来汎用計算機に依存し、そのため、処理速度・コストの点で問題があり、利用範囲は研究室内に限られていた。最近の LSI 技術の急速な進展により、低価格の分析器、合成器等が開発され、広く通信・情報処理システムにおいて実用化されるのも遠くないと思われる。

最後に、日頃、ご指導いただいている野田基礎研究部長、斎藤特別研究室長、小池第四研究室長に深謝いたします。

参 考 文 献

- 1) J. L. Flanagan: *Speech Analysis, Synthesis and Perception*, 2nd edition, Springer-Verlag, (1972).
- 2) D. R. Reddy, ed.: *Speech Recognition*, Academic Press, (1975) Pt. 2.
- 3) M. R. Schroeder: "Vocoders: Analysis and Synthesis of Speech," *Proc. IEEE*, 54, 5, p. 720 (May 1966).
- 4) 大泉充郎監修, 藤村靖編: *音声科学*, 東大出版会 (1972).
- 5) 中田和男: *音声*, 日本音響学会編音響工学講座 7 (1977).
- 6) F. J. Harris: "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform", *Proc. IEEE*, 66, 1, p. 51 (1978).
- 7) A. M. Noll: "Cepstrum Pitch Determination", *J. Acoust. Soc. Am.*, 41, p. 293 (Feb. 1967).
- 8) R. W. Schafer and L. R. Rabiner: "System for Automatic Formant Analysis of Voiced Speech", *J. Acoust. Soc. Am.*, 47, 2, p. 634 (1970).
- 9) D. R. Reddy: "Computer Recognition of Connected Speech", *J. Acoust. Soc. Amer.*, 42, 2, p. 329 (1967).
- 10) B. Gold and L. R. Rabiner: "Parallel Processing Techniques for Estimating Pitch Periods of Speech in the Time Domain", *J. Acoust. Soc. Amer.*, 46, 2, p. 442 (1969).
- 11) G. G. Bell et al.: "Reduction of Speech Spectra by Analysis-by-Synthesis Techniques", *J. Acoust. Soc. of Amer.* 33, p. 1725 (1961).
- 12) 藤崎, 佐藤: 各種フォルマント抽出方式におけるフォルマント周波数抽出誤差の比較, *日本音響学会音声研究委員会資料 S 73-20* (1973-12).
- 13) F. Itakura and S. Saito: "Analysis Synthesis Telephony based on the Maximum Likelihood Method", *Reports of 6th Int. Cong. Acoust.*, ed. Y. Kohashi, Tokyo, C-5-5 (1968).
- 14) J. D. Markel and A. H. Gray: *Linear prediction of Speech*, Springer-Verlag, (1976).
- 15) J. Makhoul: "Linear Prediction: A Tutorial Review", *Proc. IEEE* 63, p. 561 (1975).
- 16) B. S. Atal and S. L. Hanauer: "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave", *J. Acoust. Soc. Amer.*, p. 637 (1971).
- 17) 板倉, 斎藤: 偏自己相関係数による音声分析合成系, *日本音響学会講演論文集*, 昭 44. 10, p. 199 (1971).
- 18) J. Makhoul: "New Lattice Methods for Linear Prediction", 1976 *IEEE Int. Conf. Acoust. Speech. Signal Proc.*, p. 462 (April 1976).
- 19) N. J. Miller: "Pitch detection by data reduction", *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-23, pp. 72~79 (Feb. 1975).
- 20) N. C. Geçkinli and D. Kavuz: "Algorithm for Pitch Extraction Using Zero-Crossing Interval Sequence", *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-25, pp. 559~564 (Dec. 1977).
- 21) M. M. Sondhi: "New methods of pitch extraction", *IEEE Trans. Audio Electroacoust.*, Vol. AU-16, pp. 262~266 (June 1968).
- 22) J. J. Dubnowski, R. W. Schafer, and L. R. Rabiner: "Real-time digital hardware pitch detector", *IEEE Trans. Acoust., Speech, and Signal Processing*, Vol. ASSP-24, pp. 2~8 (Feb. 1976).
- 23) L. R. Rabiner: "On the use of autocorrelation analysis for pitch detection", *IEEE Trans. Acoust., Speech, and Signal Processing*, Vol. ASSP-25, pp. 24~23 (Feb. 1977).
- 24) M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg, and H. J. Manley: "Average magnitude difference function pitch extractor", *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-22, pp. 353~362 (Oct. 1974).
- 25) C. K. Un and S. Yang: "A pitch extraction

- algorithm based on LPC inverse filtering and AMDF", IEEE Trans. Acoust., Speech, and Signal Processing, Vol. ASSP-25, pp. 565~572 (Dec. 1977).
- 26) 北脇, 橋本: 直接形 PARCOR 分析におけるピッチ抽出の検討, 日本音響学会講演論文集 2-4-1, 1975-10.
- 27) J. D. Markel: "The SIFT algorithm for fundamental frequency estimation", IEEE Trans. Audio Electroacoust., Vol. AU-20, pp. 367~377 (Dec. 1972).
- 28) M. R. Schroeder: "Period histogram and product spectrum: new methods for fundamental-frequency measurement", J. Acoust. Soc. Amer., Vol. 43, No. 4, pp. 829~834 (April 1968).
- 29) 猪股: 電子計算機による新しい音声の基本周期抽出法, 日本音響学会誌, 第16巻, 第4号(1960).
- 30) L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, and C. A. McGonegal: "A comparative performance study of several pitch detection algorithms", IEEE Trans. Acoust., Speech, Signal, Proc., Vol. ASSP-24, pp. 399~418 (Oct. 1976).
- 31) 小林, 山本: 残差信号による電話音声のピッチ抽出, 日本音響学会講演論文集 1978-5.
- 32) 東倉, 橋本: 駆動音源信号成分の高精度な抽出による合成音の品質改善, 日本音響学会講演論文集 1-4-5, 1975-10.
- 33) B. S. Atal and L. R. Rabiner: "Pattern recognition approach to voiced-unvoiced-silence classification", IEEE Trans. Acoust., Speech, Signal Proc. Vol. ASSP-24 (June 1976).
- 34) D. L. Cohn and J. L. Melsa: "The residual encoder—an improved ADPCM system for speech digitization", IEEE Int. Conf. Comm., 30~26 (June 1975).
- 35) S. H. Qureshi and G. D. Forney, Jr.: "A 9.6/16 kb/s speech digitizer", IEEE Int. Conf. Comm., 30~31 (June 1975).
- 36) B. S. Atal, M. R. Schroeder, and V. Stover: "Voice-excited predictive coding system for low bit-rate transmission of Speech", IEEE Int. Conf. Comm., 30~37 (June 1975).
- 37) A. H. Gray, Jr., R. M. Gray, and J. D. Markel: "Comparison of optimal quantizations of speech reflection coefficients", IEEE Acoust., Speech, Signal Proc. Vol. ASSP-25, pp. 9~23 (Feb. 1977).
- 38) 電子通信学会編: 聴覚と音声, 1966.
- 39) 北脇, 板倉, 斎藤: PARCOR 形音声分析合成系における最適符号構成, 電子通信学会論文誌, Vol. 61A, No. 2, 1978年2月.
- 40) 北脇, 板倉: PARCOR 係数の非線形量子化と不均一標本化による音声の能率的符号化, 電子通信学会論文誌 Vol. 61A, No. 6 (1978-6).
- 41) J. Max: "Quantizing for minimum distortion", IRE Trans. IT-6, pp. 7-12 (1960).
- 42) 伊藤, 斎藤: 線形予測残差信号における個人差について, 日本音響学会講演論文集, 1-4-6, (1975-10).
- 43) 東倉, 板倉: PARCOR 帯域圧縮方式における音声品質向上, 電子通信学会論文誌, Vol. 61A, No. 3 (1978-3).

(昭和53年3月16日受付)

(昭和53年5月8日再受付)