

## ■2 群 (画像・音・言語) - 8 編 (音声・オーディオ符号化)

# 1 章 音声・オーディオ符号化の基本技術

(執筆者：守谷健弘) [2009 年 12 月 受領]

### ■概要■

多様な音声・オーディオ符号化方式も元をたどれば 1960 年から基礎研究が始まったデジタル信号処理の基本方式が土台になっている。このうち本節では音声が発声する信号であり、また音声・オーディオ信号がほとんど人間の耳で聞くことを前提に作られているといった特殊性に関連する基本技術を中心に紹介する。

### 【本章の構成】

本章では、1-1 節で音声・オーディオ符号化方式の分類と要素技術について解説する。さらに、1-2 節で多くの方式で共通に使われる基本的な要素技術を紹介し、1-3 節で基本方式を紹介する。これらの基本方式は、各種の標準化方式の土台となっているものがほとんどである。

## ■2 群 - 8 編 - 1 章

### 1-1 音声・オーディオ符号化方式の分類と要素技術

(執筆者：小澤一範) [2009年12月 受領]

音声・オーディオ符号化は、おおまかな分類を図 1・1 に示すように、音声波形を忠実に量子化して伝送する波形符号化、音源と声道からなる音声合成モデルのパラメータを入力音声から抽出し、伝送し受信側でモデルに従い合成するスペクトル符号化、波形符号化とスペクトル符号化の両者の利点を取り入れ音声合成モデルを有し音源モデル等のパラメータは波形領域の距離尺度を用いて決定するハイブリッド符号化に大別される<sup>1,2)</sup>。

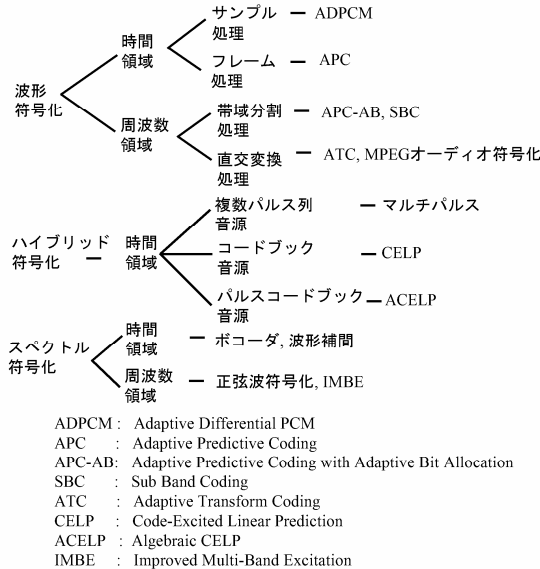


図 1・1 音声・オーディオ符号化の分類と基本方式

波形符号化は、ビットレートが 10 kbit/s 以上であり、時間領域の冗長性に着目した時間領域処理と周波数領域の冗長性に着目した周波数領域処理にわけられる。さらに、前者はサンプル処理とフレーム処理にわけられ、サンプル処理の代表には ADPCM が、フレーム処理の代表には APC が提案されている。

周波数領域処理は、帯域分割処理と直交変換処理にわけられ、前者には APC-AB が、後者には ATC や MPEG オーディオ符号化が提案されている。

スペクトル符号化は、ビットレートは 4.8 kbit/s 以下と圧縮度は高い。時間領域と周波数領域にわけられ、前者にはポコーダや波形補間符号化が、後者には正弦波符号化や IMBE があげられる。

ハイブリッド符号化は、ビットレートは 4.8～16 kbit/s が多い。励振音源により分類され、複数パルス列音源を用いるマルチパルス符号化、コードブック（符号帳）音源を用いる CELP、

パルスコードブック音源を用いる ACELP に大別される。

音声符号化に共通で用いられる要素技術としては、音声のスペクトル包絡を線形予測分析により少ない次数のスペクトルパラメータで効率的に表す PARCOR や LSP があり、これらは波形符号化の APC および APC-AB、ハイブリッド符号化の全ての方式、スペクトル符号化のボコーダと波形補間符号化で用いられている。

一方、オーディオ符号化に共通の要素技術としては、MDCT、QMF や聴覚モデルがあり、MPEG オーディオ符号化で使われている。

以降では、これらの要素技術を 1-2 節で、基本方式を 1-3 節で、それぞれ説明する。

#### ■参考文献

- 1) 守谷健弘, “音声符号化,” 電子情報通信学会, 1998.
- 2) 小澤一範, “音声符号化最新技術と応用/実現手法,” トリケップス, 1998.

## ■2群-8編-1章

### 1-2 要素技術

(執筆者：守谷健弘) [2009年12月受領]

1960年代から音声のデジタル処理の基礎研究が開始され、なかでも線形予測分析<sup>1,2)</sup>の技術は1970年代の低ビットの音声ボコーダの実現に大きく貢献した。それにとどまらず、PARCOR<sup>3)</sup>やLSP<sup>4)</sup>は、1990年代の携帯電話などに不可欠の技術として世界中で使われている。オーディオの符号化では、周波数領域の処理が大きな役割を果たし、特にMDCTやQMFが大きな役割を果たしてきている。また、聴覚特性を利用することで品質の低下を抑えつつ情報圧縮効果を高めている。

#### 1-2-1 PARCOR (Partial AutoCorrelation)

PARCOR (偏自己相関係数) は、全極型の線形予測係数と相互に変換可能なパラメータの組の1つであり、1969年に電電公社 (現NTT) の板倉、斎藤らにより提案された<sup>3)</sup>。

予測係数を係数とする多項式を分母とするシステム (全極、再帰) を仮定するが、この係数を高次と低次を反転させて直交化するときの正規化係数が本パラメータである。分析法は、短時間音声信号の自己相関係数から逐次計算により線形予測係数及び偏自己相関係数を求める方法 (Durbin-Levinson-Itakura 法)、PARCOR の縦続接続による方法 (格子法)、両者の中間的な方法 (変形格子法) などがある。

すべての係数の絶対値が1以下であれば、対応する全極型のシステムは安定である。このような安定判別の容易さや、このパラメータをそのまま使った格子形フィルタで合成と分析ができるので、線形予測係数の量子化、適応フィルタのハードウェア化に有効である。断面積が変化する音響管の反射係数<sup>5)</sup>と対応づけることもできる。この係数は音声出力用のボコーダ、通信用波形符号化 (VSELP など)、ロスレス符号化などにも使われている。

図1・2は全極型の合成フィルタ、図1・3はPARCOR係数による合成フィルタを示す。

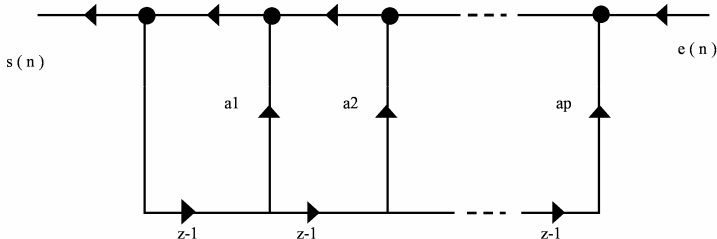


図1・2 全極型線形予測合成フィルタ

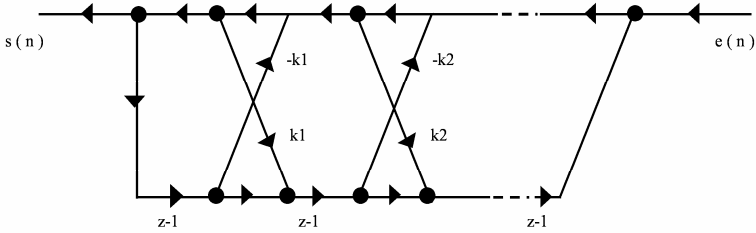


図 1-3 PARCOR 係数による線形予測合成フィルタ

### 1-2-2 LSP (Line Spectrum Pair)

線スペクトル対と呼ばれ、全極型の線形予測係数と相互に変換可能なパラメータの組の 1 つである。PARCOR よりさらに情報圧縮可能なパラメータとして、1975 年に電電公社（現 NTT）の板倉らにより提案された<sup>4)</sup>。

予測係数を係数とする多項式を分母とするシステム（全極、再帰）を仮定するが、多項式を偶対称と奇対称の 2 つの多項式の和で表現し、それぞれの Z 平面の単位円上の根の偏角が本パラメータである。分析法には、線形予測係数を求めた後に代数方程式の根を計算する方法や、係数の DFT などにより求める方法がある。

LSP は、係数の順序により安定性の判別ができること、量子化特性や補間特性等が優れ、パラメータの変動に対するスペクトル包絡の変動に対する感度がほぼ一定であることなどの特徴がある。このため、低ビットレートのボコーダだけでなく携帯電話向けの標準音声符号化方式など、大多数の音声符号化の予測係数の量子化に使われている。

図 1-4 は LSP パラメータをそのまま使った合成フィルタの構成である。

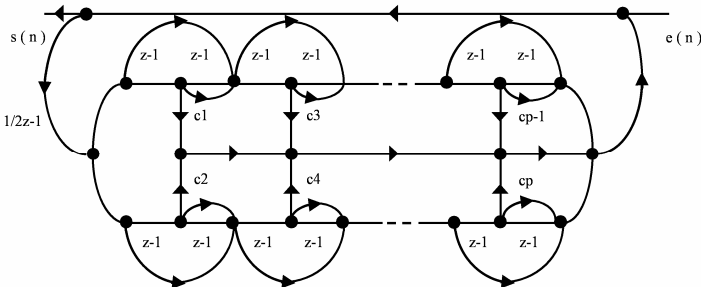


図 1-4 LSP パラメータによる線形予測合成フィルタ

### 1-2-3 MDCT (Modified Discrete Cosine Transform)

オーディオの符号化には MDCT (Modified Discrete Cosine Transform: 変形離散余弦変換) が頻繁に使われる。1986 年に AT&T ベル研究所の J.H. Princen らにより考案された<sup>6)</sup>。時間領域でオーバーラップがあるためにフレーム境界の歪が目立たない。周波数領域の独立なサン

ブル点は時間領域のサンプル点と同数であるという点が、特に波形の歪を小さくする符号化に適している。変換基底ベクトルの前半の  $N$  点が奇対称、後半の  $N$  点が偶対称で、時間領域の  $N$  サンプルの波形のうち、奇対称な成分（自由度  $N/2$ ）は後半の変換、偶対称な成分（自由度は  $N/2$ ）は前半の変換に含まれる。見かけ上  $2N$  点の時間領域のサンプルにまたがる変換であるが、自由度は  $N$  点で、周波数領域でも  $N$  点のサンプルが得られる。

通常の定常的なオーディオ信号では、 $N$  を大きくすることで周波数分解能をあげ、ビット割り当てによる変換効率を高めることができる。ただし、時間領域の振幅の変動が激しい場合は  $N$  を大きくすると時間領域の歪が問題となる。周波数領域の歪最小化によって、時間領域のフレーム内の振幅が小さい部分に対しても均等に量子化歪が発生するので、プリエコーという顕著な歪として聞こえてしまうからである。

これを防ぐ有力な方法は、時間領域での変動が激しい場合だけは、例外的に  $N$  を通常の  $1/6$  から  $1/8$  に小さくすることである。この場合、短い窓と通常の窓で遷移する MDCT 係数を設計しておく必要がある。図 1・5 は  $N = 16$  の場合の MDCT の変換係数を表示したものである。

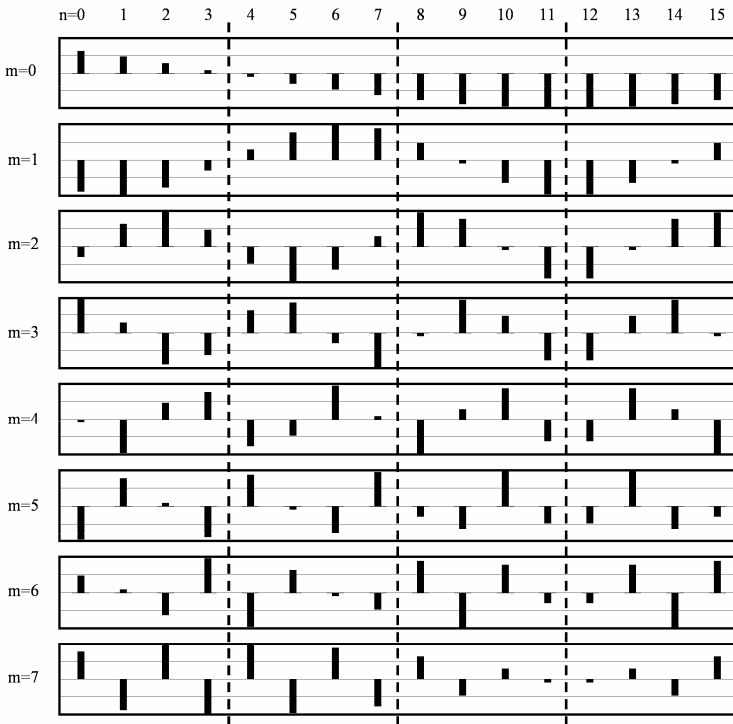


図 1・5 MDCT 変換係数の例 ( $N = 16$  の場合)

### 1-2-4 QMF (Quadrature Mirror Filter)

帯域の分割と合成には QMF (Quadrature Mirror Filter: 直交鏡像フィルタ) 形式<sup>7)</sup> のフィルタバンクが頻繁に使われる。

QMF の基本は、帯域の 2 分割フィルタでローパスフィルタ出力 (低域) と、それに対応するハイパスフィルタ出力 (高域) がペアとなる。低域成分は、ローパスフィルタ出力をそのまま 2 点に 1 点だけのダウンサンプルし、高域成分はフィルタ係数に 1 と -1 を交互に畳み込んだものをダウンサンプルする。高域信号は、また 1 と -1 を交互に畳み込んだ補間フィルタ (元のローパスフィルタと同じ)、低域は、元と同じ補間フィルタで、それぞれ 2 倍のサンプル数の系列を再構成し、それを加え合わせると完全に元の全帯域信号に戻る。高域と低域の境界での互いのエイリアシングの影響を気にしなければ、フィルタ係数は非常に短いものでも済ませることができる。1 と -1 を畳み込むのではなく、複素単位円上の等間隔に配置される N 種類の係数を畳み込み、それぞれを N 点ごとに間引いて取り出すことで N 帯域の分割信号を得ることができる。これは、位相が異なる正弦波で周波数領域に変調することでもあり、poly-phase (多相) QMF フィルタバンクとも呼ばれる。

逆に、共役の複素数を畳み込んで加え合わせることで、元の信号を再構成できる。時間領域と周波数領域の中間の状態、時間分解能と周波数分解の適度な組み合わせが選択できるといってもよい。帯域ごとのエネルギーの偏りや周波数分解能の観点では分割数が多いほうがよく、その極限は DCT や MDCT になる。

分割数が少ない場合には、再サンプルしたあとでも時間領域の波形としての連続性やサンプル間の相関が残り予測が可能であり、時間分解能を維持することが可能である。

QMF のフィルタバンクは MPEG-1, MPEG-2 のオーディオ符号化, MPEG サラウンド, SBR (Spectral Band Replication: スペクトル帯域複製) の処理のベースとなっている。

### 1-2-5 聴覚モデル

オーディオの符号化では、周波数領域のマスクング特性が符号化の聴感改善に使われている。周波数領域のマスクングは、ある周波数の信号が存在すると隣接する周波数の信号の一定レベル以下の信号が耳には聞こえなくなる現象である<sup>8)</sup>。

マスクングの影響が及ぶ帯域の幅を臨界帯域といい、低周波数領域で狭く、周波数が高くなるにつれて拡大する。臨界帯域幅に比例する非線型な周波数軸の伸縮目盛りがバーク尺度である。量子化などの処理もこのバーク尺度に比例する区切りで分割することが多い。

オーディオ信号の符号化では、臨界帯域幅ごとの入力のエネルギーを元にマスクングで聞こえなくなるレベルを計算し、量子化の雑音はそのレベル以下になるように最適な雑音制御を行う。結果的には、入力信号の振幅が大きいところに大きな雑音を許すことになる。聴覚特性に関係なく物理的な量子化雑音を最小化すると、全帯域に渡って量子化雑音が均一になることと対比すると、全体で同じビット数でも、相当大きな聴感歪削減効果がある。

#### ■参考文献

- 1) 斎藤収三, 板倉文忠, “音声スペクトル密度の統計的最適識別法に関する理論的考察,” 音響誌, vol.23, p.181, 1967.
- 2) B.S. Atal and M.R. Schroeder, “Predictive coding of speech signals,” IEEE Proc. Conf. Commun. Process., pp.

360-361, 1967.

- 3) 板倉文忠, 斎藤収三, “偏自己相関係数による音声分析合成,” 日本音響学会講演論文集, pp.199-200, 1969.
- 4) 板倉文忠, “線形予測係数の線スペクトル表現,” 日本音響学会音声研資, S75-34, 1975.
- 5) F. Itakura and S. Saito, “Analysis synthesis telephony based on the maximum likelihood method,” Report of the 6th Int. Cong. Acoust., C-5-5, 1968.
- 6) J. Princen, J. Johnston, and A. Bradley, “Subband/transform coding using filter bank designs based on time domain alias cancellation,” IEEE Proc. ICASSP, pp.2161-2164, 1987.
- 7) D. Esteban and C. Galand, “Application of quadrature mirror filters to split band voice coding schemes,” IEEE Proc. ICASSP, pp.191-194, 1977.
- 8) 三浦種敏 監修, “新版 聴覚と音声,” 電子情報通信学会 編, 1980.



## ■2群-8編-1章

### 1-3 基本方式

(執筆著者：小澤一範) [2009年12月 受領]

本節では、1-1 節の図 1・1 に示した分類において音声符号化の基本的な方式の概要を説明する。スペクトル符号化からは、まずボコーダを説明し、近年研究が盛んになっている正弦波符号化、MELP、IMBE を説明する。波形符号化からは、時間領域の符号化として APC と APC-AB を周波数領域の符号化として SBC と ATC を取り上げる。ハイブリッド符号化では、マルチパルス符号化において導入された合成による分析 (A-b-S: Analysis-by-Synthesis) 手法により音源モデルのパラメータ (ここではマルチパルスの位置と振幅) を算出する枠組みを説明する。

次に、この枠組みをもとに励振信号を符号帳 (コードブック) により符号化することで符号化効率のさらなる改善をはかった、CELP 符号化を説明する。さらに、CELP 符号化において、励振信号を代数構造のパルスコードブックにより符号化し、高速探索法を導入することでコードブック探索の演算量を大幅に削減するとともに、コードブックを格納するためのメモリ量を不要とした ACELP 符号化を説明する。なお、オーディオ符号化については 2-5 節と 3-3 節で説明を行う。

#### 1-3-1 ボコーダ

(執筆著者：小澤一範) [2009年12月 受領]

AT&T ベル研究所の H. Dudley らにより考案されたボコーダ<sup>1)</sup>がルーツである。その後、チャンネルボコーダ、ホルマントボコーダなどいくつかの方式が研究されたが<sup>2)</sup>、線形予測分析 (Linear Predictive Coding: LPC)<sup>3,4)</sup> に基づく線形予測ボコーダ<sup>5,6)</sup> の提案により研究が加速された。

ボコーダは、音源モデルと声道モデルから構成される音声合成モデルを用いる。音源モデルは、有声の場合はインパルス発生器からの出力で、無声の場合は雑音発生器からの出力で表わされる。声道モデルは線形予測ボコーダでは全極型フィルタ (All-Pole Filter) により構成される。線形予測ボコーダの受信側のブロック図を 1・6 に示す。

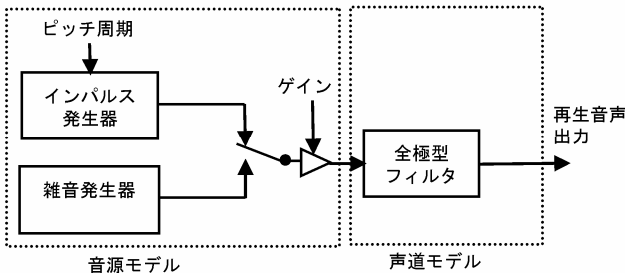


図 1・6 線形予測ボコーダの受信側の構成

分析合成に用いる場合、入力音声から短時間ごとに線形予測分析法を用いて全極型フィルタの係数を計算し、有声か無声かの判別や、有声の場合はピッチ周期、音源のゲインを抽出

し、一定区間ごとに伝送する。ここで、線形予測係数をそのまま量子化して伝送すると比較的多くのビット数が必要であるだけでなく、伝送路誤りや受信側での補間などによりフィルタの安定性が保証されない恐れがあるという問題が明らかにされ、PARCOR 係数<sup>7)</sup>あるいは K パラメータ<sup>6)</sup>に変換することで量子化ビット数の削減とフィルタ安定性の確認ができることが解明された。その後、さらに量子化感度の低い LSP<sup>8)</sup>が提案され、さらなる低ビットレート化が可能になった。音質面では、明瞭性は高いものの自然性や話者ごとの声質の違いの再現性は十分ではない。

### 1-3-2 波形補間

(執筆者：小澤一範) [2009年12月 受領]

スペクトル符号化の枠組みの中で、ポコーダの改善として 1995 年に W. Kleijn らにより提案された<sup>9)</sup>。音声波形を周期波形と非周期波形に分離し、周期波形はピッチ周期の波形を補間することにより効率的に符号化する。

時間的にゆっくり変化する成分 (SEW) と変化が早い成分 (REW) の和で特性波形を表し、送信側で音声信号に線形予測分析を行なって求めた予測残差波形からこれらの情報を抽出して伝送する。受信側では、SEW は時間方向で補間し REW はランダムな位相を与えることにより、特性波形を復元し、線形予測合成フィルタにより音声を再生する。本符号化を用いて音声を 2.4 kbit/s で符号化した事例が報告されている。

### 1-3-3 正弦波符号化

(執筆者：小澤一範) [2009年12月 受領]

音声波形を複数個の正弦波の重ね合わせによる正弦波モデルで表し、正弦波の振幅、周波数、位相を伝送する方式で、1986 年に MIT の R. McAuley らにより提案された<sup>10)</sup>。音声の低ビットレート符号化に適用するために、正弦波の振幅スペクトルの包絡を 10~20 次の線形予測モデルで近似することにより、パラメータを減らす方法が提案されている。この方法によれば、フレーム長を 30 msec、合成時は 7.5 msec ごとにパラメータを補間しながら合成することにより、2.4 kbit/s で 4.8 kbit/s CELP とほぼ同等の音質が得られると報告されている。

なお、正弦波の周波数を音声のピッチ調波構造の周波数に一致させた方式は、ハーモニック符号化 (Harmonic Coding) と呼ばれる。

### 1-3-4 APC (Adaptive Predictive Coding)

(執筆者：小澤一範) [2009年12月 受領]

音声信号の予測符号化に線形予測分析を適用する先駆的な方式であり、1967 年に AT&T ベル研究所の B. Atal, M. Schroeder らにより提案された<sup>4)</sup>。送信側のブロック図を図 1-7 に示す。適応予測器として、近接相関に基づく短期予測とピッチの長期相関に基づく長期予測の両者を併用し、これらを用いて音声信号を予測して得られた予測残差信号を適応量子化することを特徴とする。

ここで、短期予測器の予測係数は短時間音声信号から線形予測分析法 (LPC) を用いて計算され、長期予測器の係数は短期予測残差信号から計算される。短期予測器は 16 次、長期予測器は 3 次とし、短期予測器および長期予測器の両者を用いて音声信号を予測して得られた予測残差信号を適応量子化器により量子化し、量子化された予測残差信号、両予測器の係数と適応量子化器の量子化ステップサイズを伝送する。予測残差信号は 8 kbit/s で、両予測器の係数は合計で 3~5 kbit/s で伝送することができる<sup>11)</sup>。さらに、ノイズシェイピングの手法を

導入することで音質改善をはかることができる<sup>12)</sup>。これは、短区間での量子化雑音の周波数特性を聴覚のマスキング効果を利用して聴感的にマスクさせるように変形させることにより実現する。適応予測符号化や量子化雑音のマスキングは、後に提案されるマルチパルス符号化<sup>13)</sup>や CELP 符号化<sup>14)</sup>の基礎をなす枠組みと位置づけられる。またノイズシェイピングの考え方は、ボコーダ駆動 ATC<sup>15)</sup>やオーディオ符号化に取り入れられている。

適応予測符号化に後述のサブバンド符号化を組み合わせ音質改善する方法として、電電公社(現 NTT)の菅田、板倉らにより適応ビット割り当て適応予測符号化 (APC-AB) が提案された<sup>16)</sup>。本方法は、音声信号を帯域分割し、信号パワの周波数方向の偏り及び時間方向の偏りを利用して適応的なビット配分を行ない、各帯域において割り当てられたビット数で適応予測符号化を行なう。

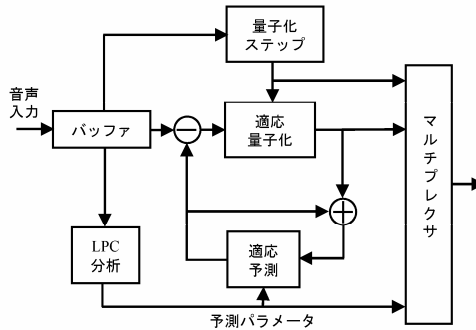


図 1-7 APC 符号化の送信側のブロック図

### 1-3-5 ATC (Adaptive Transform Coding)

(執筆者：守谷健弘) [2009年12月 受領]

適応変換符号化はブロック符号化の代表例の1つで、ブロック単位の波形を周波数領域に変換して量子化する。1977年にゲッチンゲン大学の P. Noll らにより提案されたものは<sup>17)</sup>、スペクトル包絡の形状を対数スペクトルの折れ線近似で補助情報として伝送し、フレームごとに各 DCT 成分に対する最適量子化ビット数を計算してスカラ量子化することを基本にしている。また、フレーム境界で接続歪みが生じるので台形窓を使ってフレーム境界を連続的につないでいる。

上記の ATC は特に入力信号の統計的特徴に依存しないため、音声の符号化に限れば量子化雑音はそれほど小さくできない。これに対し VD-ATC (Vocoder Driven ATC: ボコーダ駆動 ATC)<sup>15)</sup> は入力を音声信号に限定して、16 kbit/s 以下の低ビット化をねらったものである。特にスペクトル包絡を線形予測分析で求め、予測パラメータを量子化する点とピッチ周期を求めて予測を行う点が線形予測ボコーダと共通する。

以上は音声信号を直接変換するものであるが、線形予測残差信号を変換して量子化する方法も提案されている。このときは、各成分の分散は一定と近似できるが、最終的に合成フィルタの周波数特性(パワスペクトラム)を考慮すると、上記と等価な情報割り当てに帰着される。また、逆変換後に線形予測合成フィルタを通すのでフレーム境界歪みが小さいという利点もある。さらに、係数をベクトル量子化することで歪みを小さくする方法が提案されて

いる。この場合には、適応情報割り当てに従って数多くのビット数に対応できるベクトル量子化が必要である。

さらに、楽音の符号化法として比較的変換長の大きい MDCT と高度な聴覚心理モデルを用いた ASPEC が典型的な変換符号化として知られている。このほか、音楽、オーディオの符号化のほとんどは適応変換符号化に基づくものである。

### 1-3-6 SBC (Sub Band Coding)

(執筆者：守谷健弘) [2009 年 12 月 受領]

サブバンド符号化は、AT&T ベル研究所の R. Crochiere らにより提案された<sup>18)</sup>。音声信号を周波数方向で複数の帯域に分割して符号化する方式で、帯域分割符号化とも呼ばれる。帯域ごとにダウンサンプルして符号化するので、量子化サンプル数は帯域分割しない場合と同じであるが、分割帯域内ではスペクトルのダイナミックレンジが小さく予測等の処理が楽になる。また、量子化雑音は帯域外にでることはなく、符号器あるいは復号器に必要な帯域の選択が容易にできる。

例えば、情報が少ないときには全帯域で量子化歪みを発生させるのではなく、高域を送らず低域のみを量子化することにより量子化歪みを小さくするとか、伝送路が混んでいるときには復号器側で高域の情報を捨てるといった処理が可能である。さらにハードウェア構成上、演算語長が少なくてすむとか、並列処理が容易といった利点もある。一方、直交変換符号化と同様に、サブバンド符号化でも帯域ごとにそのエネルギーの対数に比例させて、適応的に情報配分を行わないと全帯域の量子化歪みを小さくすることができない。これに対応させて数多くの種類の情報量の量子化器をもつことは一般にわずらわしく、システムを複雑にするので、平均的なパワースペクトルに合わせて固定の情報配分を行う場合も多い。

### 1-3-7 マルチパルス符号化

(執筆者：小澤一範) [2009 年 12 月 受領]

マルチパルス励振線形予測符号化と呼ばれる。線形予測ボコーダや低ビットレートでの適応予測符号化の音質を改善し、自然で良好な音声を再生することを目的として、1982 年に AT&T ベル研究所の B. Atal らにより提案された<sup>19)</sup>。図 1・8 に原理図を示すように、特徴は振幅が非零の複数個のパルス列から励振音源信号が構成されること、合成による分析の考えを取り入れ、送信側に励振音源信号と線形予測分析により構成された全極型フィルタからなる音声合成部をもち、これによる合成音声を一旦計算し、短時間入力音声信号との誤差信号の電力を最小化するように、励振音源信号のパラメータ（パルスの位置と振幅）を求めることである。

さらに、聴感的な音質を改善するために、誤差信号を聴感的な尺度に対応する聴感重みづけフィルタに通した上で誤差電力を最小化する。送信側ではフレーム (10~20 msec) ごとに全極型フィルタの線形予測係数を計算し、サブフレーム (5~10 msec) ごとにパルスの位置と振幅を計算して伝送する。受信側では送信側と同じ音声合成部により音声を再生する。

この枠組みにより、音声を 10 kbit/s で伝送しても自然な音質が得られるようになった。信号を合成しながらパルスの位置と振幅を計算するのは演算量的に膨大になるため、全極型フィルタと重みづけフィルタの縦続接続からなるフィルタのインパルス応答の自己相関と、入力信号を重みづけフィルタに通した重みづけ信号とインパルス応答との相互相関とを用いて、1 パルスずつ逐次計算することで、実用的な演算量におさえる方法が提案された<sup>19)</sup>。これに

より DSP (Digital Signal Processor) などを用いて実時間演算が可能となり、有線伝送装置にも適用された。さらに、長期予測器によるピッチ予測の導入<sup>20, 21)</sup>、パルス振幅の同時最適化手法<sup>22)</sup>や、ベクトル量子化の導入<sup>23)</sup>などの種々の改良方式が提案され、さらなるビットレート削減を可能にした。

なお、後述の符号励振線形予測符号化 (CELP)<sup>24)</sup> は、マルチパルス符号化の基本的な枠組みを踏襲し<sup>25)</sup>、励振音源信号やフィルタ係数の量子化などにベクトル量子化の手法を取り入れ、符号化効率をさらに改善している。

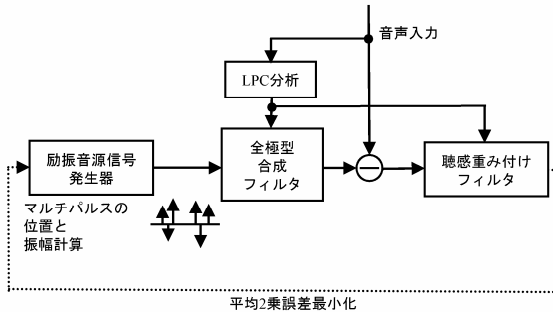


図 1-8 マルチパルス符号化の原理

### 1-3-8 CELP (Code-Excited Linear Prediction)

(執筆者：小澤一範) [2009年12月 受領]

符号励振線形予測符号化と呼ばれ、1984年に AT&T ベル研究所の B. Atal と M. Schroeder らにより提案された<sup>26)</sup>。短区間 (5 msec) 音声信号の線形予測残差信号がガウス分布を有するランダム信号で近似できることから、マルチパルス符号化の枠組みにベクトル量子化の考えを導入し、駆動音源信号を 10 ビット、40 次元のガウス乱数コードブック (符号帳) でおきかえた<sup>25, 26)</sup>。原理図を図 1-9 に示すように、送信側ではガウス乱数をピッチ予測器と線形予測分析による LPC 合成フィルタに通して再生した音声と入力音声との間で、5 msec ごとに重みづけ誤差電力を計算し、これを最小化する最適なコードベクトルを選択する。選択されたコードベクトルを表す符号と、ピッチ遅延、ピッチゲイン、フィルタ係数を短時間ごとに伝送する。本構成により、音源信号を 2 kbit/s で伝送することが可能であることを示した。

当初の提案ではコードブックを全探索していたので、Cray-1 スーパーコンピュータを使用しても 1 sec の音声を符号化するのに 125 sec かかっていたが<sup>26)</sup>、その後、多岐にわたる研究が集中的に行なわれ、演算量の大幅な削減とビットレート低減化が行われ、実用化を可能にした。

主な技術としては、コードブックの構造化<sup>27)</sup>、コードブックの高速探索手法<sup>28)</sup>、特にピッチ周期がサブフレーム長よりも短い話者の音質を改善するための適応コードブック<sup>29)</sup>、非整数遅延<sup>30)</sup>、LSP 係数のベクトル量子化によるビットレート削減<sup>31, 32)</sup>、受信側のみの処理で量子化雑音をマスクして音質を改善する、ポストフィルタ<sup>33)</sup>やピッチプリフィルタなどがあげられる。これらにより、CELP を基盤技術とした改良方式が 1989 年以降に TIA、電波産業会 (ARIB)、ITU-T、ETSI、3GPP や 3GPP2 などで、固定網やモバイル網向けに続々と標準化され、携帯電話や IP 電話などのグローバルな普及に貢献した。

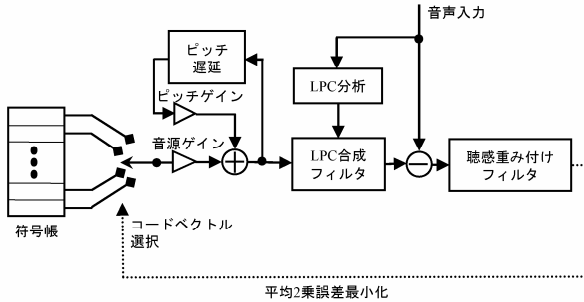


図 1・9 CELP 符号化の原理

### 1-3-9 ACELP (Algebraic CELP)

(執筆者：小澤一範) [2009年12月 受領]

代数符号励振線形予測符号化と呼ばれ、1987年にカナダ・シェルブルーク大学の J. Adoulらにより提案された<sup>34)</sup>。振幅を $\pm 1$ とした少数の個数のパルスからなる代数符号コードブック (Sparse Algebraic Codebook) 構造に特徴がある。

図 1・10 に示すように、取りえる位置をパルスごとに重複なく、複数のトラックごとに等間隔にあらかじめ制約して決めておき、これらから 1 つを選択し、複数のトラックからのパルスを重ね合わせる。各パルスの振幅の絶対値が 1 であること、パルスの個数が少数であることを利用し、時間軸逆向きフィルタリング (Backward Filtering) を導入した高速探索法により、歪みを最小化するパルスの位置を高速に探索することができる。本構造によれば、規則によりパルスを発生させるのでコードブックを格納するメモリが不要であり、高速探索法とあいまって、携帯電話などで使用される低電力汎用 DSP でも CELP を実現することができるようになった。

さらに、ACELP をベースとする改良方式が有線電話向けの ITU-T G.729<sup>35)</sup> などや、携帯電話向けの AMR<sup>36)</sup> などの標準方式に取り入れられ、普及に貢献した。なお、有線電話向けの標準方式は第 2 章で、携帯電話向けの標準方式は第 3 章でそれぞれ説明する。

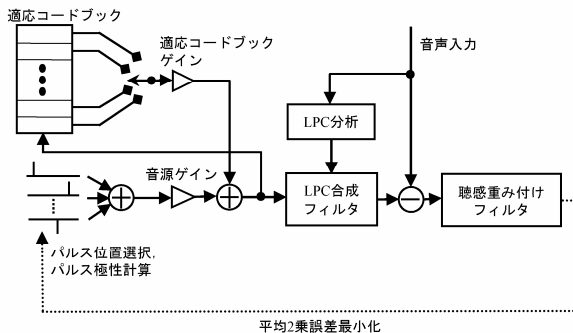


図 1・10 ACELP 符号化

### 1-3-10 MELP (Mixed Excitation LP)

(執筆: 芹沢昌宏) [2009年12月 受領]

線形予測ポコーダを基本とする方式で、Texas Instruments 社で開発された。1997年に米国国防省 (Department of Defence: DoD) の Federal Standard として、ビットレート 2.4 kbit/s、フレーム長 22.5 msec の MELP<sup>37)</sup> が採用され、軍事通信、衛星通信や緊急通信向けに広く利用されている (本編 2 章 2-4-3 参照)。

従来の線形予測ポコーダがピッチ励振信号を用いたのに対して、図 1-11 に受信側のブロック図を示すように混合励振 (Mixed-Excitation) 信号を用いることにより、ブザー音などを低減している。更に、周期・非周期インパルス、適応スペクトル強調、パルス拡散フィルタなどの採用によってより自然な音声を合成できる。実雑音環境でも高い性能を示し、従来の DoD FS1016 4.8 kbit/s ポコーダを越える音質を達成している。

また、MELP の拡張としてビットレートを半減した 1.2 kbit/s MELPe (enhanced-MELP) も開発されており、2001年に DoD で採用されている。

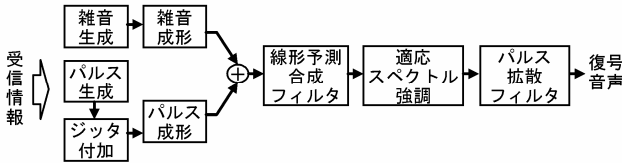


図 1-11 MELP の受信側のブロック図

### 1-3-11 IMBE (Improved Multi-Band Excitation)

(執筆: 守谷健弘) [2009年12月 受領]

IMBE<sup>38)</sup> は INMARSAT (INternational MARitime SATellite: 国際海事衛星) の M システムと呼ばれる国際船舶電話用途として 1990年に標準化された 4.15 kbit/s の符号化方式である。

MIT の J. Lim 教授らの提案であり、その基本は MIT で 1980 年ころからの正弦波モデル (sinusoidal model) あるいはハーモニック符号化 (harmonic coding) にある。すなわち、音声波形をピッチ周波数の整数倍 (ハーモニクス) の正弦波の重ね合わせで合成する。

### 1-3-12 広帯域オーディオ符号化

(執筆: 守谷健弘) [2009年12月 受領]

48 kHz サンプルを基本とするオーディオ符号化は、図 1-12 のような周波数領域の符号化が主体である。時間軸でオーバーラップのある MDCT 係数を適応的に量子化することで、情報圧縮を実現する。適応的な量子化の際には聴覚のマスキング特性を考慮し、可変長符号化やベクトル量子化が使われる。適応量子化に必要な MDCT 係数の概形 (スペクトル包絡) は別途補助情報として符号化される<sup>39, 40)</sup>。この基本枠組みで、インターネットや携帯プレーヤで広く使われている MP3 (MPEG-1,2 LayerIII)、デジタル放送や携帯プレーヤに使われている AAC (Advanced Audio Coding) が制定された。詳しくは第 2 章を参照されたい。

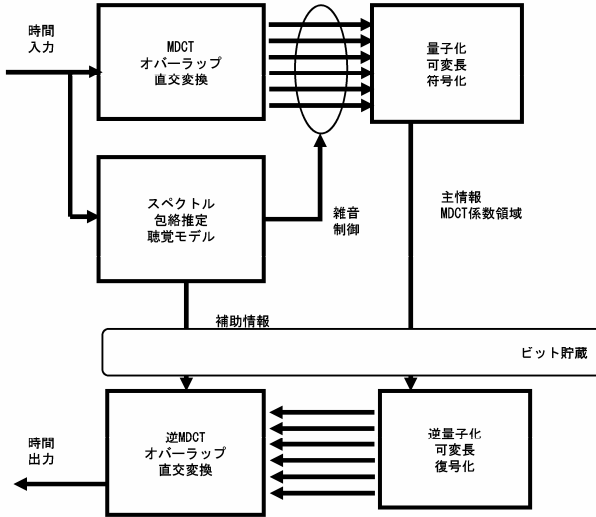


図 1・12 周波数領域のオーディオ符号化

#### ■参考文献

- 1) H. Dudley, "The vocoder," Bell Labs Record, vol.18, no.4, pp.122-126, 1939.
- 2) M.R. Schroeder, "Vocoders: Analysis and synthesis of speech," Proc. IEEE, vol.54, no.5, pp.720-734, 1966.
- 3) 斎藤収三, 板倉文忠, "音声スペクトル密度の統計的最適識別法に関する理論的考察," 音響誌, vol.23, p.181, 1967.
- 4) B.S. Atal and M.R. Schroeder, "Predictive coding of speech signals," IEEE Proc. Conf. Commun. Process., pp. 360-361, 1967.
- 5) F. Itakura and S. Saito, "Analysis synthesis telephony based on the maximum likelihood method," Report of the 6th Int. Cong. Acoust., C-5-5, 1968.
- 6) B.S. Atal and S.L. Hanauer, "Speech analysis and synthesis by linear prediction," J. Acoust. Soc. Am., vol.50, pp.637-655, 1971.
- 7) 板倉文忠, 斎藤収三, "偏自己相関係数による音声分析合成," 日本音響学会講演論文集, pp.199-200, 1969.
- 8) 板倉文忠, "線形予測係数の線スペクトル表現," 日本音響学会音声研資, S75-34, 1975.
- 9) W. Kleijn and J. Haagen, "Waveform interpolation for coding and synthesis," Chapter 5 in Speech coding and synthesis, Elsevier, 1995.
- 10) R.J. McAuley and T.F. Quatieri, "Speech analysis-synthesis based on a sinusoidal representation," IEEE Trans. Acoust. Speech and Signal Process., ASSP-34, pp.744-754, 1986.
- 11) B.S. Atal and M.R. Schroeder, "Predictive coding of speech signals," Bell System Technical J., pp.1973-1986, 1970.
- 12) B.S. Atal and M.R. Schroeder, "Predictive coding of speech signals and subjective error criteria", IEEE Trans. Acoust. Speech and Signal Process., ASSP-27, no.3, pp.247-254, 1979.
- 13) B.S. Atal and J.R. Remde, "A new model of LPC excitation for producing natural sounding speech at low bit rates," IEEE Proc. ICASSP, pp.614-617, 1982.
- 14) B.S. Atal and M.R. Schroeder, "Stochastic coding of speech signals at very low bit rates," IEEE Proc. ICC, pp.1610-1613, 1984.



- 15) J.M. Tribolet and R.E. Crochiere, "Frequency domain coding of speech," IEEE Trans. Acoust. Speech and Signal Process., ASSP-27, pp.512-530, 1979.
- 16) M. Honda and F. Itakura, "Bit allocation in time and frequency domains for predictive coding of speech," IEEE Trans. Acoust. Speech and Signal Process., ASSP-32, pp.465-473, 1984.
- 17) R. Zelinski and P. Noll, "Adaptive transform coding of speech signals," IEEE Trans. Acoust. Speech and Signal Process., ASSP-25, pp.299-309, 1977.
- 18) R.E. Crochiere, S.A. Webber, and J.L. Flanagan, "Digital coding speech in sub-bands," Bell System Technical J., pp.1069-1085, 1976.
- 19) 小澤一範, 小野茂, 荒閑卓, "マルチパルス駆動型音声符号化法の検討," 信学技報, CS82-16, 1983.
- 20) 小澤一範, 荒閑卓, "ピッチ情報を用いたマルチパルス駆動型音声符号化法," 音響学会秋季講演, 2-2-14, 1983.
- 21) S. Singhal and B.S. Atal, "Improving performance of multi-pulse LPC coders at low bit rates," IEEE Proc. ICASSP, pp.1.3.1-1.3.4, 1984.
- 22) K. Ozawa, S. Ono, and T. Araseki, "A Study on pulse search algorithms for multi-pulse excited speech coder realization," IEEE JSAC, vol.SAC-4, no.1, pp.133-141, 1986.
- 23) 小澤一範, 田海真一, 野村俊之, "マルチパルスベクトル量子化音源と高速探索に基づく MP-CELP 音声符号化," 信学論, vol.J79-A, no.10, pp.1655-1663, 1996.
- 24) B.S. Atal and M.R. Schroeder, "Stochastic coding of speech signals at very low bit rates," IEEE Proc. ICC, pp.1610-1613, 1984.
- 25) B.S. Atal, "High-quality speech at low bit rates: Multi-pulse and stochastically excited linear predictive coders," IEEE Proc. ICASSP, pp.1681-1684, 1986.
- 26) M.R. Schroeder and B.S. Atal, "Code-excited linear prediction (CELP): High quality speech at very low bit rates," IEEE Proc. ICASSP, pp.937-940, 1985.
- 27) T. Moriya, "Two-channel conjugate vector quantizer for noisy channel speech coding," IEEE JSAC, vol.10, no.5, pp.866-874, 1992.
- 28) G. Davidson and A. Gersho, "Complexity reduction methods for vector excitation coding," IEEE Proc. ICASSP, pp.3055-3058, 1986.
- 29) W.B. Kleijn, D.J. Krasinski, and R.H. Ketchum, "Improved speech quality and efficient vector quantization in SELP," IEEE Proc. ICASSP, pp.155-158, 1988.
- 30) P. Kroon and B.S. Atal, "Pitch predictors with high temporal resolution," IEEE Proc. ICASSP, pp.661-664, 1990.
- 31) 管村昇, 板倉文忠, "線スペクトル対 (LSP) 音声分析合成方式による音声情報圧縮," 信学論 A, vol.J64-A, pp.599-606, 1981.
- 32) 大室仲, 守谷健弘, 間野一則, 三樹聡, "移動平均型フレーム予測を用いる LSP パラメータのベクトル量子化," 信学誌, vol.J77-A, no.3, pp.303-313, 1994.
- 33) J.-H. Chen and A. Gersho, "Adaptive postfiltering for quality enhancement of coded speech," IEEE Trans. Speech and Audio Proc., vol.3, no.1, pp.59-71, 1995.
- 34) J.P. Adoul, P. Mabilieu, M. Delprat, and S. Morissette, "Fast CELP coding based on algebraic code," IEEE Proc. ICASSP, pp.1957-1960, 1987.
- 35) R. Salami, C. Laflamme, J.P. Adoul, A. Kataoka, S. Hayashi, T. Moriya, C. Lamblin, D. Massaloux, S. Proust, P. Kroon, and Y. Shoham, "Design and description of CS-ACELP: Toll quality 8 kb/s speech coder," IEEE Trans. Speech and Audio Processing, vol.6, no.2, pp.116-130, 1998.
- 36) 3GPP TS26.090, "AMR speech codec: Transcoding functions," 1999.
- 37) A. McCree, K. Truong, E.B. Georg, T. Barnwell, and V. Viswanathan, "A 2.4 kbit/s MELP coder candidate for the new US Federal Standards," IEEE Proc. ICASSP, pp.200-203, 1996.
- 38) J.C. Hardwick and J.S. Lim, "The application of the IMBE speech coder to mobile communications," IEEE Proc. ICASSP, pp.249-252, 1991.
- 39) M. Bosi and R.E. Goldberg, "Introduction to digital audio coding and standards," Kluwer Academic Publisher, 2003.
- 40) 北脇信彦 編著, "ディジタル音声・オーディオ技術," 電気通信協会, 未来ねっと技術シリーズ, 1999.