

音響におけるデジタル技術入門

山崎 芳 男

ま え が き

最近、好むと好まざるとにかかわらずコンピュータがなんらかの形で我々の日常生活にかかわりのある存在になってきました。一昔前には考えられなかったことですが、LSI技術の進歩に伴いマイクロプロセッサという形で家庭にまでCPUが侵入しつつあります。

音響の分野においてもデジタル技術は従来一部制御系に使われていましたが、最近では信号系へも盛んに導入されつつあります。これは電子計算機による処理と専用ハードウェアによる処理に分けられ、電子計算機による処理は音声信号の分析、合成の分野では比較的早くから使われており、ハードウェアによる処理は電話回線のPCM多重伝送や宇宙通信に見られるように通信の分野で乗用に供されています。

しかし広帯域音響信号を扱うオーディオの分野では、周波数帯域やダイナミックレンジの上でAD、DA変換器や記憶素子の容量等に制約があり、デジタル技術が導入されたのはごく最近のことです。実用化されているものでは英国BBCのPCM中継網、日本コロムビアのPCMレコーダ、スタジオ用ディレマシシ等が上げられます。特に従来のアナログ伝送系で問題の多かったテープレコーダについては、研究用、市販を目的としたもの等が次々と発表されています。

ところで、信号のデジタル伝送の特長は“0”、“1”の判断が正確に出来る伝送路を使いさえすれば信号劣化がなく、伝送特性が基本的には標本化周波数と量子化ステップ数によってのみ決まるという点にあります。さらにデジタルでは常に安定した正確な処理が行えるので、精度上アナログ処理では実現不可能であった微妙な処理も可能です。またデジタル処理の各種機能は集積回路化しやすいので、信頼性の向上、小形化、省電力、価格低減が期待できます。

山崎芳男：早稲田大学理工学部

以上のようにデジタル処理には幾多の利点がありますが、一般にアナログ伝送、処理に比較して広い伝送帯域を要し、また伝送系に誤りがあると検出訂正のために装置も大がかりにならざるを得ません。従って、現段階では音響信号のデジタル処理はある程度採算を度外視しても性能を確保する意味のある分野で実用に供されているといえましょう。

このたび本誌から、「音響におけるデジタル技術入門」という題で3回の連載記事の御依頼を受けました。そこで音響信号のそれも広帯域音響信号の伝送、記録、処理へのデジタル技術の導入について、

第1回 デジタル技術の基礎

第2回 音響測定へのデジタル技術の導入

第3回 音響機器へのデジタル技術の導入

の内容で話を進めさせていただくことにします。多少なりとも読書諸兄のお役に立てば幸いです。

1. デジタル伝送

デジタル化とは連続量 - アナログ量 - を離散的な値で表現すること、あるいはアナログ量のうち離散的な点の値にのみ意味を持たせる操作のことである。

図1にデジタル伝送の基本構成を示す。振幅が時間の関数である信号の時間方向のデジタル化を行うのが標本化であり、振幅をデジタル化するのが量子化である。量子化した信号をパルス符号化して伝送するのがPCM伝送である。

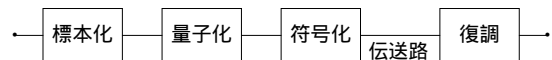


図1 デジタル伝送系の基本構成

1.1 標本化

標本化とは連続信号のある時点の値を読み取る操作である。帯域制限を受けた時間関数を表現するには一定の離散的間隔で標本化を行えば十分である。すなわち、標本化定理によれば、

「帯域制限された時間信号を帯域の少なくとも2倍の周波数で標準化すれば、その標準列に原信号の全情報が含まれる。」

からである。いま $\frac{1}{2T}$ (Hz) に帯域制限された信号を $x(t)$ とすると、 $x(t)$ は、標本 $x(nT)$ を使って、

$$x(t) = \sum_{n=-\infty}^{+\infty} x(nT) \cdot \frac{\sin \frac{\pi}{T}(t-nT)}{\pi(t-nT)} \quad \dots (1)$$

かける。(1)式は原信号 $x(t)$ が帯域 $\frac{1}{2T}$ (Hz) の理想

帯域フィルタに全標本を通すことにより再現されることを示している。

現実には完全な帯域制限、幅のないパルスによる標本化、幅のないパルス列と理想(低域)フィルタによる復調は実現不可能である。

帯域制限が不完全であると標本化した信号のスペクトルに重なりが生ずる。標本化による一度生じた重なりは分離不可能である。この重なりによるひずみを折り返しひずみと呼ぶ。

一方、有限幅パルスで復調を行うと保持効果により周波数伝送特性が変化する。いま標本化周期を T (秒)、パルス幅を τ (秒) とすると、純音 $x(t) = \cos 2\pi ft$ に対する復調信号 $y(t)$ は、

$$y(t) = \frac{1}{T} \cdot \frac{\sin \tau \cdot \pi \cdot f}{\pi \cdot f} \cdot \cos 2\pi ft \left(t - \frac{\tau}{2} \right) \quad \dots \dots \dots (2)$$

となる。これが保持効果で時間遅れは波形にひずみを与えないが、振幅周波数特性は図2のように変化する。

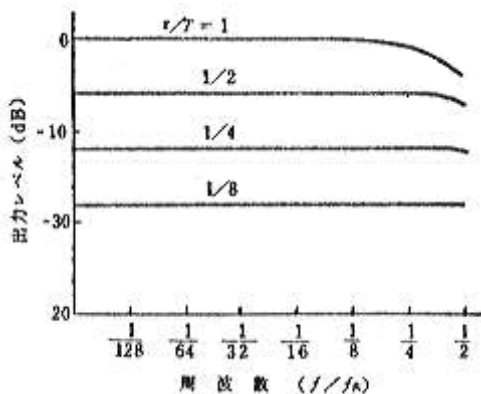


図2 保持効果

ところで広帯域音響信号の伝送には、帯域は20kHz程度必要である。従って、標準化周波数は40kHz以上

必要である。一方、保持効果については、 $\frac{\tau}{T}$ をあまり小さくすると S/N が悪化するので、振幅周波数特性との兼ね合いから $\frac{\tau}{T} = \frac{1}{4} \sim \frac{1}{10}$ が適当といえよう。

1.2 量子化

標本値を離散的ないくつかの振幅値で表現するのが量子化である。そして、この振幅値は一般には2進符号を使って表現される。2進数の1桁が1ビット(BIT, Binary Digit)とよばれ、8ビットでは $2^8 = 256$ 、12ビットでは $2^{12} = 4096$ 段階の振幅値が表現できる。

ところで量子化は連続する振幅を離散的な値で表現する操作であるから、丸めによる誤差を生ずる。この誤差を量子化雑音とよび、PCMが原理的に有する雑音である(図3)。

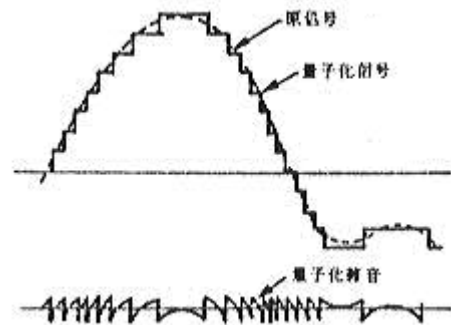


図3 量子化雑音

一様量子化による雑音電力 N_q は、量子化ステップを ΔV とすると、

$$N_q = \frac{(\Delta V)^2}{12} \text{ (W)} \quad \dots \dots \dots (3)$$

として与えられる。従って、最大振幅の正弦波信号と量子化雑音の S/N_q は、ビット数を M とすると、

$$S/N_q = \frac{3}{2} \cdot 2^M = 6M + 1.8 \text{ (dB)} \quad \dots \dots \dots (4)$$

となる。ビット数と量子化雑音の関係を図4に示す。

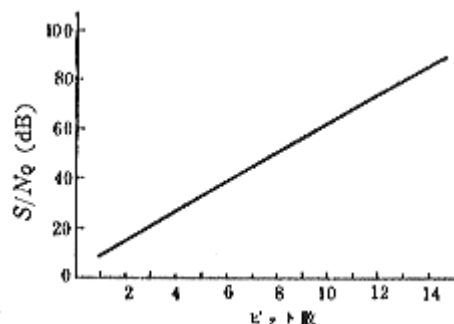


図4 一様量子化の S/N_q

広帯域音響信号の伝送においてダイナミックレンジは70 ~ 80dBは必要である。従って、一様量子化では少なくとも12 ~ 13ビットを要する。

量子化雑音は半導体の接合部雑音や熱雑音とは異なり、信号の存在するときだけ発生する一種のひずみである。

量子化ビット数の多い場合そのスペクトルは白色に近づく。最下位ビット (LSB) にかからない程度の小振幅の信号は、大きなひずみ (granular distortion) を発生する。このひずみを低減する目的で “dither” とよばれる小振幅の白色雑音や正弦波を付加することがある。BBC では周波数が標準化周波数の 1/2 で、振幅が量子化ステップの 1/2 の方形波と、元来の量子化雑音レベル + 4dB に相当する白色雑音を重畳して聴感上 “granular distortion” の影響を低減している⁽¹⁾。

1.3 非線形量子化

一樣量子化においては S/N_0 は信号の振幅に比例する。振幅の小さい信号に対しては量子化ステップを小さく、振幅の大きい信号に対しては量子化ステップを大きくとり (圧縮)、復調時に逆の操作 (伸張) を行えば相対的に小振幅時の S/N_0 は改善される。また、大振幅時の悪化をある程度許容すれば量子化ビット数が節約できる。圧伸特性はデジタル段で論理演算により与える場合とアナログ段で非線形回路素子により与える場合とがある。

ところで、アナログ伝送において、テープレコーダのダイナミックレンジ拡大に用いられるドルビーや dBx 等のノイズリダクション装置も一種の圧伸装置である。これらのアナログ圧伸装置ではレベル検出はある時定数をもたせて行うが、デジタル伝送では一般に各標本ごとにレベルに応じた圧伸を行う。これを瞬時圧伸とよび各標本毎に仮数情報 M ビットと指数情報 L ビットを

伝送することになるが、BBC 等から指数情報を N 標本に一度伝送する準瞬時圧伸の実験結果が報告されている⁽²⁾。準瞬時圧伸では標準列を N 標本ずつのブロックに分割し、ブロック内の標準の最大値または電力により量子化ステップを決定する。従って、各標本に平均した伝送ビット数は $(L/N+M)$ ビットとなり、瞬時圧伸に比較して大幅に減少する (図 5)。

各種の楽音について標準化周波数 48kHz, 12 ビット一様最量化と遜色のないビット数を求めると、瞬時圧伸では 10.8 ビット、準瞬時圧伸では 9.7 ビット程度であると報告されている⁽²⁾。

非線形量子化によるレベル圧伸は、電話等の狭帯域伝送系で低レベル時の S/N_0 を改善して聴感上の条件を良くする場合には有効であるが、広帯域音響信号伝送においては 1 ~ 2 ビットの節減しか期待できない。

後述の予測符号化、エントロピー符号化等による伝送はビット節減効果も大きく無ひずみ伝送が可能であるから、伝送路の節約という観点からははるかに有利である。

広帯域音響信号伝送においては非線形量子化は補助的手段とし、通常のレベルに対しては一樣量子化を行い、ごくまれに生ずる大振幅時のみ減衰器を動作させ等価量子化ステップを大きくして過負荷レベルを引き上げ、ダイナミックレンジを拡大する目的で使用するのが有効である。仮数部は 9 ビットあれば楽音に対しては十分であるから⁽³⁾、仮数部を 10 ビット、指数部を 3 ビットとすると 100dB 以上のダイナミックレンジと 60dB 以上の

2. 符号化

2.1 情報量とエントロピー

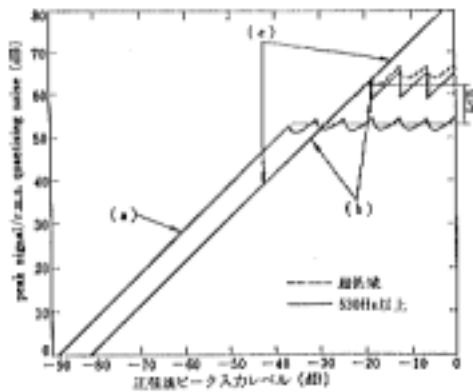
よく犬が人を噛んでもニュースにはならないが人が犬を噛めばニュースになるなどという。これは犬が人を噛むのは当たり前だが人が犬を噛むのはめずらしいからである。これを情報量という観点から見ると犬が人を噛むのは情報量が小さく、人が犬を噛むのは情報量が大きということになる。すなわち情報量は確率の少ないめずらしい事象ほど大きく、確率に対して単調減少関数であり、その事象の起こる確率を p とすると、

$$\text{情報量} = -\log_2 p \text{ (bit)}$$

と定義される。

例えばトランプの 1 枚をひいてハートを取る確率は 1/4 であるから、ハートをひいたという事象の情報量は、

$$-\log_2 \frac{1}{4} = 2 \text{ (bit)}$$



- (a) 瞬時圧伸7セグメント10ビット伝送
- (b) Near-instantaneous 4-segment discontinuous low 10.07ビット伝送
- (c) 一樣符号化13ビット伝送

図 5 BBC の Near-instantaneous 4-segment discontinuous low による S/N

である。

ところで、確かにめずらしい事象が起きたことを知れば、その情報量は大きい、珍しいことは当然めつたに起きない。それ程めずらしくはないが多少変わったことが起きたということをそのたびに知れば、長い時間観察した総情報量はかえって大きくなる可能性もある。このような観点から見た、長い間の平均的な情報量をエントロピーと呼ぶ。

いま、例えばある1年の毎日の天気確率が晴 p_1 、曇 p_2 、雨 p_3 であったとすると(ただし $p_1 + p_2 + p_3 = 1$ とする。),

確率	情報量	日数	1年の情報量
晴 p_1	$-\log_2 p_1$	$p_1 \times 365$	$-365 p_1 \log_2 p_1$
曇 p_2	$-\log_2 p_2$	$p_2 \times 365$	$-365 p_2 \log_2 p_2$
雨 p_3	$-\log_2 p_3$	$p_3 \times 365$	$-365 p_3 \log_2 p_3$

となるから、全情報量は $365(p_1 \log_2 p_1 + p_2 \log_2 p_2 + p_3 \log_2 p_3)$ となる。従ってエントロピー(平均情報量) H は、 $H = -p_1 \log_2 p_1 + p_2 \log_2 p_2 + p_3 \log_2 p_3$ 、一般形で書くと、

$$H = -\sum_i p_i \log_2 p_i \text{ (bit)}$$

となる。

2.2 マルコフ過程のエントロピー

先に述べたように、我々が対象とする音響信号は時間とともに変化する統計的な事象である。このような時間の関数としての確率的な事象を確率過程という。これに対し時間によらず一定な事象は決定的な事象であり確率過程とはいわない。

ところで確率過程にもいろいろあるが、音楽信号のようにある時点の信号が過去の信号の影響を受けるような確率過程をマルコフ過程という。さらにある時点の事象が過去の 個の事象のみの影響を受けるときこれを 重マルコフ過程という。普通、一重マルコフ過程を単純マルコフ過程と称する。

これに対して信号の各時点の事象が確率的に独立である情報源を“記憶のない情報源”と呼ぶことがある。

マルコフ過程のエントロピーは前節で述べた単純な生起確率から定義される平均情報量とは違って、各々の状態における遷移確率の平均となる。すなわち、マルコフ過程では、

$$H = -\sum_{i,j} p(i)p(j|i) \log p(j|i) \text{ (bit)}$$

となる。ここで $p(j|i)$ は i から j への遷移確率で、 $\sum_j p(j|i) = 1$ である。

例えばPCM伝送において、“0”、“1”の生起確率が各々1/2であったとしよう。“0”、“1”の生起が独立と

考え生起確率からエントロピーを計算すると、

$$H = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1 \text{ (bit)}$$

となる。

次に一重(単純)マルコフ過程として観察したら、“0”からは7/8が“1”へ遷移し、1/8が“0”留まり、“1”からは3/4が“0”へ遷移し、1/4が“1”に留まるものとしよう。このとき、

$$H = -\frac{1}{2} \left(\frac{7}{8} \log_2 \frac{7}{8} + \frac{1}{8} \log_2 \frac{1}{8} \right) - \frac{1}{2} \left(\frac{3}{4} \log_2 \frac{3}{4} + \frac{1}{4} \log_2 \frac{1}{4} \right) = 0.677 \text{ (bit)}$$

となる。

これはたとえ“0”、“1”の生起確率が1/2であってもマルコフ過程においてはエントロピーが低下することを示している。

2.3 エントロピーと冗長度

前の例ではエントロピーの最大値は $p(0) = p(1) = 1/2$ のときで $H = 1$ である。 n 種の情報について一般形で書くと、最大エントロピー H_{\max} は、

$$H_{\max} = H \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n} \right) = \log_2 n$$

となる。すなわち、各事象が独立でかつ等しい生起確率を持つときエントロピーは最大となり、普通はこれ以下の値をとる。このエントロピーの最大値 H_{\max} と実際のエントロピーの比を1から差引いた値 $1 - \frac{H}{H_{\max}}$ をその信号の冗長度という。

前節の例では、

$$\text{冗長度} = 1 - \frac{H}{H_{\max}} = 1 - 0.677 = 0.323 \text{ (bit)}$$

となる。

ところで、誤りのある伝送路を使って正確に情報を送るには、誤りに見合った冗長度が必要である。しかし、冗長度があれば正確に伝わるわけではなく、その冗長度を利用した適切な誤り検出、訂正機能を設けなくてはならない。

誤りの検出、訂正には情報そのものの冗長性を利用する方法と、誤り検出、訂正用の符号を別に付加する方法とがある。

2.4 符号化定理

音響信号のPCM伝送系では標本化、量子化した信号をそのまま2進符号で伝送することが多い。しかし、いままで述べたように、実際の情報源、つまり楽音等のもつエントロピーは量子化ビット数よりかなり小さい値を示す。

シャノンが情報源のエントロピー H と伝送路の符号容量 C の間に美しい定理を導いている。それは伝送路に誤りのない場合の第1定理と伝送路に誤りがある場合の第2定理である。

第1定理

情報源のエントロピーを H , 伝送路の符号容量を C とすると,

- (1) この情報源から単位時間当たり (C/H) 個以上の情報源は送れない。
- (2) (C/H) 個以下で送る符号は必ず存在する。

第2定理

情報源のエントロピーを H , 伝送路に誤りが存在しその符号容量を C とすると,

- (1) $H \leq C$ ならば誤り率をいくらでも0に近づける符号化法が存在する。
- (2) $H > C$ のときは誤り率0の符号化法は存在しない。
- (3) $H > C$ ならば, あいまい度を $H - C$ にいくらでも近づける符号化法が存在する。

符号化については第3回目にふれる予定であるが, 一つだけ計算例を示す。我々はテープレコーダの伝送特性に問題があるとしてかなり大がかりなPCMレコーダを作ったりするが, テープレコーダの音も一度の録音再生にはほぼ十分の音質をもっている。そこでテープレコーダの伝送特性を一応帯域 20kHz, 雑音は白色で S/N 40dB と仮定して, 伝送路としてみた場合の符号容量を求めてみよう。

帯域を $W(\text{Hz})$ とし, 雑音全電力が $N(W)$ の正規分布の伝送路の単位時間当たりの符号容量 C は,

$$C = W \log_2 \frac{S+N}{N} \quad (\text{bit/sec})$$

である。従ってオーディオ用テープレコーダの帯域を 20kHz, $(S+N)/N$ を 40dB とすると $C = 267$ (kbit/s) となる。

一方, PCM系の量子化ビット数 13, 標準化周波数 40kHz とすると, 独立な信号(前述のように楽音ではずっと低いはず)を許容しても信号源の情報量 H は $13 \times 40 = 520$ (kbit/s) である。従って, シャノンの符号化定理によれば, 高々 19cm/s 程度のテープレコーダのたった2トラックを使えばPCMレコーダ(1チャンネル)が実現できる符号がそ存在するはずである。

家庭用VTRに至っては帯域は優に 2MHz, S/N も 40dB はあるから, その伝送符号容量 C は 26.7Mbit/sec もあり実に 50チャンネル分のHiFi記録が可能ことになる。しかるに, 現実には4チャンネルがやっとである。

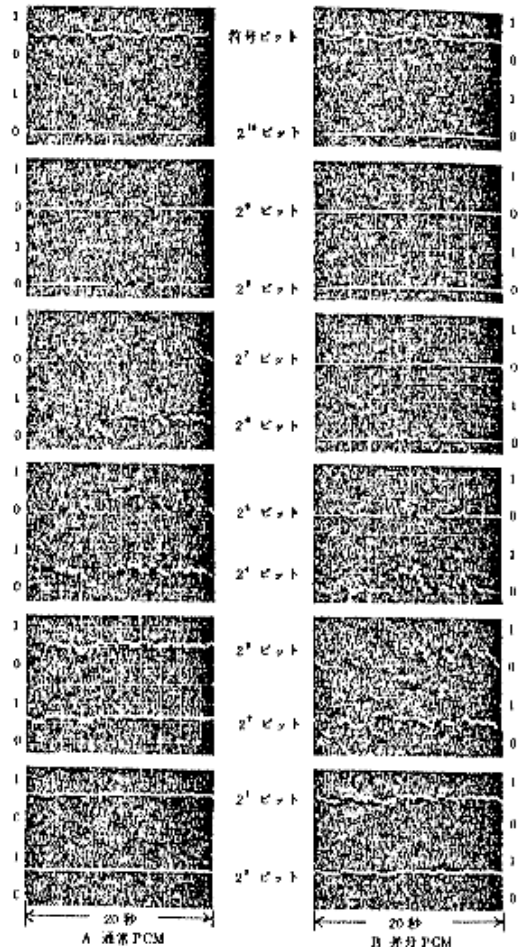


図6 各ビットの出力波形(時定数100msのCR負荷にて撮影)

まだまだ理論的には改善の余地があるということである。

2.5 符号容量とエントロピーの整合(伝送路の高度利用)

楽音(女性ボーカル)を標準化周波数 48kHz, 量子化ビット数 12, 折り返し2進符号で構成されたPCM伝送系に与えたときの各ビット波形を図6(a)に示す。また各ビットの“1”の出現確率を図7に, 6dBステップで表示したレベル分布を図8に示す, さらに信号を1,024点ハニログ窓でフーリエ変換して得られたスペクトルのピークを図9に示す⁽¹⁴⁾。

これらの結果から, 各ビットの使用状況, レベル分布にはかなり偏りがあり, スペクトル分布も高域は大幅に

時間相関およびビット間相関が存在するので、複製情報はさらに減少し、4～5ビット/標本である。従って何らかの方法で信号のもつ冗長性を除去してやれば通常PCMで12ビットの信号に何らひずみを与えることなく4～5ビットで伝送可能なはずである。

2.6 予測符号化

図6(a)の波形は楽音の時間相関が強いことを示している。予測符号化は過去の標本値から適当な方法で標本値の予測を行い、真の標本値と予測値との差を符号化する伝送方式である。標本値間の相関が強ければ強いほど符号ビット数の節減が可能である(図6(b))。ただし予測符号化伝送では復調に際し積分操作が行われるので伝送路に誤りがあると、その影響が多くの標本に及びるので、誤りの検出、訂正機構が不可欠である。

2.7 直交変換

図9に示すフーリエ変換して得られた楽音のスペクトルは高域の分布の少ないことを示している。従って、直交変換を行って統計的にレベルの低い帯域のビット数を減らすことによりビットレートの低減が可能である。直交変換としてはKarhunen・Loeve変換, Fourier変換, Hadamard変換等がある。

直交変換で注意しなくてはならないのが、切り出し窓の影響である。方形窓を用いると切り出しによる不連続により、見かけ上スペクトルが広がってしまう。一方、方形窓以外の窓を使うと波形の再現が困難になる。従って直交変換は音声信号のようにピッチという切り出し時間を決定する根拠のある信号や内像のようにフィールドの区切りのある信号には有力であるが、一般の広帯域音響信号伝送に直交変換による伝送路節約の効果はあまり期待できない。

2. エントロピー符号化

図8に示すようにレベル分布は偏りが存在する。そこで、標本値の振幅レベルの中で出現頻度の高いレベルに短い符号、出現頻度の低いレベルに長い符号を与えるエントロピー符号化を行うことにより、ビットレートを平均情報量が与える限界Rate distortion boundaryに近づけることが可能である。

このような不等長符号にはシャノンの符号、ハフマン符号等がある。先にふれた楽音(ボーカル)はハフマン符号を適用し、約3秒分ノモリーを併用することにより5ビットで無歪伝送が可能である。これはあくまで対製とした信号についてだけであるが、楽音等のレベル分布はかなり似た属性を示すので、同一符号割当てを用いて多くの楽音が5ビット程度で無歪伝送が可能である。ただし、エントロピー符号は不等長符号であるため、伝

送には大容量の記憶素子と辞書（振幅に対する符号割当て表）が必要である。従って実時間の伝送には不向きであるが、大量に取り込んだ信号を誤りの非常に少ない記憶装置に蓄えるときには大変有効な方式である。

具体的には例えば、通常の符号で取り込んで大量の信号を電子計算機に一時ファイルし、その信号の性質を調べた上でエントロピー符号を生成し、辞書（符号表）とともに恒久ファイルに蓄え、再びその内容が必要になったときに辞書に基づいて原信号を一時ファイルに復元し処理する等といった使い方に適した符号である。

3. 誤り検出，訂正符号

誤り検出，訂正符号の歴史はシャノンが、符号化第二定理「伝送路の通信容量より小さい情報を送信する場合には、誤り率を0にする符号が存在する」ことを1948年に証明した時点に始まる。以後、今日まで多くの人々が各種の符号を提唱し、そのいくつかが実用に供されている。

本稿では最終回にいくつかの具体例を示すこととし、

ここでは符号の名称と概略を示すに留める。

誤り検出符号としては、パリティチェック、CRC (Cyclic Redundancy Check) が広く実用に供されている。誤り訂正符号としては、主としてランダム誤りの訂正には巡回符号であるハミング符号やBCH符号、バースト誤り訂正にはHageIbager や岩垂 Massey の畳み込み符号が提唱されている。一般に誤り検出は比較的率よく行うことができるが、誤り訂正となるとひとすじなわけではいかないうで実用化例は少ない。

いずれにしても誤り検出，訂正にあたっては対象とする伝送路の誤りの性質を十分は握る必要がある。また、キャリア等符号そのもの以外の情報を積極的に利用するのも有効な場合がある。

例えばPCMレコーダではテープのドロップアウトに起因するバースト誤りの検出，除去に符号の他に固定ヘッド形では再生波形のエンベロープ、VTRを利用する場合にはFMのキャリアレベルを利用すると好結果が得られる。

(参考文献は次号に掲載)(以下次号)