

固体記憶素子への音響信号の記録

早稲田大学理工学部 山崎 芳 男

テープやディスクから可動部分のない非接触読み書きへ

家庭用のPCMテープレコーダやコンパクトディスクが実用化され、デジタルオーディオも定着した感がある。ところで、これらのデジタルオーディオシステムの記憶媒体は磁気テープやプレスされた円盤であり、両者とも可動部分を有している。円盤では針による接触読み出しのアナログディスクからレーザーによる非接触読み出しに変わっているが、テープでは変調方法がバイアスによる直接記録からPCMによるデジタル記録に変わっただけで、磁気ヘッドによる接触書き込み、読み出し方法には変化はない。一部でレーザー磁気記録も試みられているが、この場合も可動部分をもつことには変わりはない。

技術の動きは確実に動くものや触っているものは排除される方向にある。テープレコーダは可動部を持つうえ、テープとヘッドとの接触も不可避である。コンパクトディスクプレーヤほ現在のところすべてディス

クそのものを回転させているが、ディスクを固定してレーザー光の方を電氣的あるいは光学的に回転させる方法も考えられる。

ところで、半導体記憶素子や磁気バブル等の固体記憶素子の集積度や処理速度の向上、消費電力および価格の低減には目を見晴らせるものがある。この傾向は速度を増すことこそあれ今後も当分は続くものと思われる。従って、デジタル記憶の媒体がこれらの固体記憶素子に置き変わる日もそう遠い将来のことではないだろう。

広帯域音響信号をデジタル化するには、20～50kHzの標準化周波数と10～18の量子化ビット数が必要である。従ってデジタル化された1秒の音響信号を伝送あるいは記録するには1チャンネル当たり200～900kbitという膨大な容量を要する。64kbitの固体記憶素子を使うとすると、1秒当たり3～15個の素子が必要とする。繰り返し使用やランダムアクセスが必要な短いメッセージ等の記録、再生には現時点でも十分導入の意味があるといえるが、1分

の音響信号を記録するのに180～1,000個の素子数を必要とする。最近出回り出した256kbitの記憶素子を使っても現時点では長時間の音響信号を固体記憶素子に記憶するにはハードウェア規模や価格の面で無理がある。

将来、固体記憶素子の集積度が上がり、価格が低減すればデジタル化した音響信号をそのまま記憶するシステムが実現するものと考えられるが、固体記憶素子の符合誤りは極端に少ないので、現時点では音響信号のもつ冗長性を取り除いたうえで記録する方法が有効である。

シャノンは情報源の1秒当たりの平均情報量 S [bit/s] と伝送容量 C [bit/s] との間に“もし $S \leq C$ なる関係があれば、あいまい度をいくらでも0に近づけて伝送できる符号化方法が存在する”という定理を導いている。

デジタル化された音響信号の振幅の分布あるいはスペクトル分布にはかなりの偏りがあることは良く知られている。これらは信号の持つ冗長性に起因するものであり、信号の

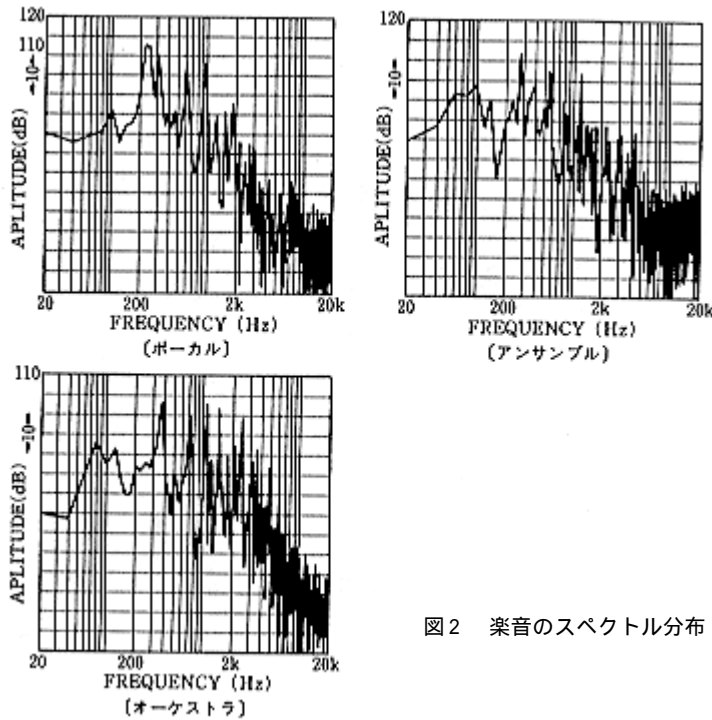


図2 楽音のスペクトル分布

ることがわかる。以上のように広帯域音響信号のデジタル化により得られた離散的な有限個の通報が等確率では発生せず、しかもそのスペクトルも全帯域に様に分布しているものではない。

エントロピーと冗長度

よく犬が人をかんでもニュースにならないが、人が犬をかめばニュースになるなどという。これは犬が人をかむのは当たり前だが、人が犬をか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、犬が人をかんだのを知るのは情報量が小さく、人が犬をかんだの知れば情報量が大きいということになる。すなわち情報量は確率の低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、
 情報量 $= -\log_2 p[\text{bit}] \dots\dots\dots (1)$
 と定義される。

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、めずらしい事象は当然めつたに起こらない。それほど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、長い間観察した総情報量はかえて大きくなる可能性もある。このような観点からみた長い間の平均的な情報量をエントロピーとよび、一般形で書くと、
 $H = \sum p_i \log_2 p_i[\text{bit}]$
 となる。

先に述べたように、われわれが対象とする音響信号は時間とともに変

化する事象である。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。確率過程にもいろいろあるが、楽音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 k 個の事象のみの影響を受けるときこれを k 重マルコフ過程と称する。これに対して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よぶ。

マルコフ過程のエントロピーは単純な生起確率から定義される平均情報量とは違って、おのおの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、

$$H = \sum p_i p(j|i) \cdot \log_2 p(j|i)[\text{bit}]$$

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

例えばデジタル伝送において、あるビットの“0”“1”の生起確率が等しく、それぞれ $1/2$ であったとしよう。“0”“1”の生起が独立と仮定して、生起確率からエントロピーを求めると、

$$H = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1[\text{bit}]$$

となる。

次に1重(単純)マルコフ過程として観察したところ“0”からは $1/8$ が“1”に遷移し、 $7/8$ が“0”に留まり、“1”からも $1/8$ が“0”に遷移し、 $7/8$ が“1”に留まっていたとしよう。この場合には、エントロピー H は

$$H = -\frac{1}{2} \left(\frac{7}{8} \log_2 \frac{7}{8} + \frac{1}{8} \log_2 \frac{1}{8} \right) - \frac{1}{2} \left(\frac{7}{8} \log_2 \frac{7}{8} - \frac{1}{8} \log_2 \frac{1}{8} \right)$$

=0.544 (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。

n 種の事象について一般形で書くと、最大エントロピー H_{max} は、

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

となる。すなわち、各事象が等しい生起確率をもつときエントロピーは最大となり、通常はこれ以下の値を示す。このエントロピーの最大値 H_{max} と実際のエントロピーとの比を1から引いた値 $1 - (H/H_{max})$ をその信号の冗長度という。

音響信号の情報量

前述のようにデジタル化された広帯域音響信号の各ビットの使用状況やレベル分布にはかなり偏りがある。さらに音楽のスペクトル構造が高域で大幅に低下している。この信号のもつ偏りは冗長度に起因するものであり、これに着目して伝送路の節約、すなわち信号のエントロピーと伝送容量の整合が可能ならずである。

各ビット間の相関はないものとして、各ビットのエントロピーを12重マルコフ過程まで計算し、その総和を求めた結果を表2に示す。ボーカルについては記憶のない確率過程として捉えると1標本当たり9.94bit、マルコフ過程として捉えると7.77bitとなる。他の音楽や道路騒音についても6~9bit/標本程度である。すなわち12bit一様量子化した信号が何ら劣化されることなく6~9bitで伝送できるはずであることを示し

ている。しかもこれはビット間の相関を無視した結果であり、実際には表1からも読み取れるように各ビット間にも強い相関関係があるので情報量も表2の結果より大幅に少ないはずである。これはデジタル化さ

れた信号を各ビットごとに独立して扱うのではなく、12bitで表現される4,096のすべての値から4,096の事象への遷移確率を調べなくてはならないということを示している。図3に同じ楽音の条件つき確率分布を

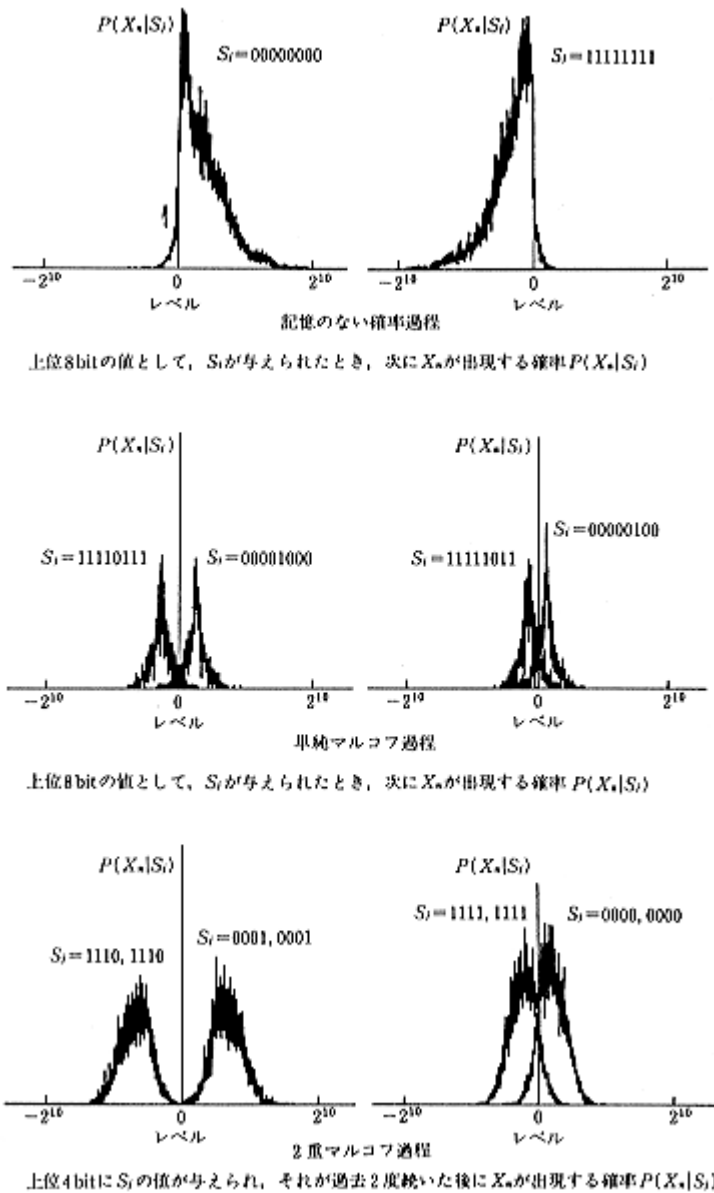


図3 楽音（オーケストラ）の条件付きレベル分布

表2 ビット間相関がないとして据えた楽音の情報量

〔ボーカル〕

(単位:ビット/標本)

BIT	独立	1重 マルコフ	2重 マルコフ	3重 マルコフ	4重 マルコフ	5重 マルコフ	6重 マルコフ	7重 マルコフ	8重 マルコフ	9重 マルコフ	10重 マルコフ	11重 マルコフ	12重 マルコフ
SIGN	1.000	0.284	0.284	0.284	0.283	0.283	0.282	0.281	0.280	0.279	0.278	0.277	0.274
2 ¹⁰	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
2 ⁹	0.261	0.090	0.090	0.089	0.088	0.087	0.086	0.085	0.085	0.083	0.076	0.075	0.075
2 ⁸	0.761	0.362	0.361	0.358	0.356	0.352	0.348	0.344	0.342	0.337	0.280	0.273	0.260
2 ⁷	0.933	0.665	0.662	0.649	0.646	0.633	0.628	0.623	0.620	0.615	0.610	0.600	0.584
2 ⁶	0.984	0.883	0.876	0.862	0.852	0.845	0.838	0.833	0.829	0.825	0.818	0.809	0.794
2 ⁵	0.997	0.970	0.966	0.961	0.956	0.952	0.948	0.944	0.941	0.938	0.907	0.925	0.912
2 ⁴	0.999	0.994	0.993	0.991	0.990	0.989	0.987	0.986	0.984	0.982	0.979	0.972	0.960
2 ³	1.000	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.998	0.996	0.994	0.988	0.977
2 ²	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.997	0.994	0.989	0.977
2 ¹	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.997	0.995	0.989	0.978
2 ⁰	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.997	0.995	0.989	0.978
合計	9.938	8.249	8.233	8.195	8.172	8.142	8.116	8.093	8.077	8.048	7.928	7.888	7.771

〔アンサンブル〕

(単位:ビット/標本)

BIT	独立	1重 マルコフ	2重 マルコフ	3重 マルコフ	4重 マルコフ	5重 マルコフ	6重 マルコフ	7重 マルコフ	8重 マルコフ	9重 マルコフ	10重 マルコフ	11重 マルコフ	12重 マルコフ
SIGN	0.998	0.373	0.372	0.372	0.371	0.371	0.371	0.371	0.370	0.369	0.368	0.367	0.366
2 ¹⁰	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
2 ⁹	0.027	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.018	0.018	0.018
2 ⁸	0.169	0.095	0.093	0.092	0.091	0.090	0.089	0.089	0.089	0.088	0.080	0.079	0.079
2 ⁷	0.606	0.332	0.330	0.327	0.324	0.320	0.318	0.316	0.315	0.313	0.271	0.267	0.261
2 ⁶	0.914	0.675	0.674	0.667	0.662	0.658	0.654	0.652	0.650	0.647	0.643	0.636	0.624
2 ⁵	0.981	0.926	0.925	0.920	0.917	0.916	0.914	0.912	0.911	0.909	0.906	0.900	0.887
2 ⁴	0.996	0.993	0.993	0.993	0.992	0.992	0.992	0.992	0.911	0.989	0.984	0.981	0.970
2 ³	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.997	0.994	0.988	0.977
2 ²	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.997	0.994	0.989	0.977
2 ¹	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.998	0.997	0.995	0.989	0.977
2 ⁰	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999	0.998	0.997	0.994	0.989	0.978
合計	8.691	7.414	7.398	7.391	7.376	7.365	7.356	7.349	7.339	7.324	7.248	7.204	7.115

〔オーケストラ〕

(単位:ビット/標本)

BIT	独立	1重 マルコフ	2重 マルコフ	3重 マルコフ	4重 マルコフ	5重 マルコフ	6重 マルコフ	7重 マルコフ	8重 マルコフ	9重 マルコフ	10重 マルコフ	11重 マルコフ	12重 マルコフ
SIGN	0.998	0.385	0.383	0.383	0.383	0.383	0.382	0.382	0.381	0.381	0.379	0.378	0.376
2 ¹⁰	0.019	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009
2 ⁹	0.601	0.199	0.199	0.199	0.197	0.196	0.194	0.192	0.190	0.188	0.162	0.161	0.159
2 ⁸	0.911	0.565	0.564	0.561	0.558	0.553	0.547	0.543	0.538	0.534	0.530	0.523	0.516
2 ⁷	0.955	0.874	0.872	0.860	0.853	0.847	0.842	0.837	0.833	0.828	0.822	0.815	0.801
2 ⁶	0.983	0.968	0.963	0.955	0.949	0.942	0.937	0.932	0.927	0.922	0.916	0.908	0.896
2 ⁵	1.000	0.993	0.988	0.980	0.971	0.962	0.954	0.947	0.942	0.937	0.933	0.926	0.914
2 ⁴	0.999	0.996	0.991	0.984	0.975	0.965	0.957	0.949	0.944	0.939	0.934	0.928	0.915
2 ³	0.995	0.992	0.988	0.982	0.974	0.967	0.958	0.952	0.945	0.941	0.936	0.929	0.917
2 ²	0.999	0.996	0.992	0.984	0.975	0.965	0.956	0.949	0.944	0.939	0.934	0.927	0.915
2 ¹	0.999	0.996	0.992	0.984	0.975	0.965	0.956	0.949	0.942	0.939	0.934	0.928	0.916
2 ⁰	0.994	0.992	0.988	0.982	0.975	0.966	0.958	0.951	0.946	0.940	0.936	0.928	0.916
合計	10.45	8.965	8.929	8.863	8.764	8.720	8.650	8.592	8.541	8.497	8.425	8.360	8.250

示す。

明らかに過去のレベルに現在の信号レベルが影響を受けており、音響信号が記憶のない情報源から発生される通報ではなく、記憶のある情報源（マルコフ情報源）から発生されるものであることがわかる。これは次のような事実から考えても明らかである。すなわち人間が音響信号、例えばレコードからの再生音を聴取している際、針飛びのクリック雑音に対し、これが雑音であるか、あるいは曲の一部であるかはすぐさま判定できる。すなわち、人間が過去信号の周波数およびレベルの変化に対して続いて到来する音を予測しつつ聴取しているということに他ならない。聴取の対象をランダム雑音とするならば、人間は次にくるべき音を予測しえない。従って音響信号が記憶のある情報源から発生されるものであると考えられる。

表3にデジタル化された信号のビット間の相関をも考慮して1重マルコフ過程として見たときの遷移確率と各ビットのエントロピーを示す。このとき各ビットのエントロピーの総和は6.523bitである。さらに2重、3重と次数を増すと、エントロピーの減少の速度は低下し、10重程度で5～6bitに収束する。

情報量と伝送容量の整合

デジタル化された音響信号にはかなりの冗長度があるので、伝送路を効率良く利用するにはエントロピーと伝送路の整合が必要なことはすでに述べた。

ところで、ビデオ信号の記録や伝送にはフレーム間あるいは走査線間の相関を利用した伝送路の節約が広く検討されていたが、音響信号については音声の伝送へのポコーダの導入等が検討されている程度で、広帯

域音響信号への導入例は殆どない。これは、帯域の広いビデオ信号をデジタル化するので、伝送路の節約を切望されているうえ、標本化周波数を決定する根拠となる同期信号やカラーのサブキャリア周波数の規定がテレビジョン信号にはすでに存在するからであろう。

なお、ここで検討する冗長度の除去は、あくまでも信号の統計的性質を根拠として行うものである。従って、一旦収録したデジタル信号を処理したうえで、能率良くファイルするような使い方には向いているが、どんな信号が到来するかわからない収音には不向きである。

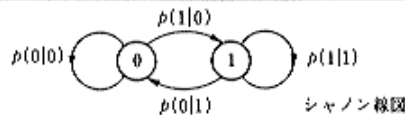
電力 P の信号を送送できる帯域 $W[H_2]$ の伝送路に電力 N の白色熱雑音が存在するとき、この伝送路の伝送容量 C は、
 $C = W \log_2(P+N)/N$ [bit/s]
 である。従って、例えば $W = 5\text{MHz}$ 、 $(P+N)/N = 36\text{dB}$ のVTRをPCM録音にすると、理論的には30Mbit/sの記録が可能になる。これは実にデジタル化された音響信号をそのまま記録しても40～60チャンネル、冗長度の除去を行えば実に150チャンネルの記録が可能であることを示している。しかるに現状の記録密度はこの1/5～1/10程度で、まだまだ高密度化の余地が残されているといえよう。

予測符号化

予測符号化は過去のいくつかの標本値 x の線形結合として現在の標本の推定値 x_0 を求め、真の標本値と予測値 x_0 との差である予測誤差 $e = x_0 - x_e = x_0 - \sum ai xi$ を符号化して伝送するものである。

表3 マルコフ過程として据えた楽音の遷移確率と情報量 (弦楽四重奏)

ビット	$p(0 0)$	$p(1 0)$	$p(0 1)$	$p(1 1)$	$H(\text{ビット})$
符号	0.9418	0.0582	0.0612	0.9388	0.3148
2^{16}	0.9999	0.0001	0.3189	0.6811	0.0000
2^9	0.9987	0.0013	0.1749	0.8251	0.0048
2^8	0.9959	0.0041	0.0898	0.9102	0.0558
2^7	0.9822	0.0178	0.1347	0.8653	0.1805
2^6	0.9411	0.0589	0.2239	0.7761	0.4157
2^5	0.8487	0.1513	0.3421	0.6579	0.7093
2^4	0.7355	0.2645	0.4199	0.5801	0.8907
2^3	0.6405	0.3595	0.4527	0.5473	0.9650
2^2	0.5762	0.4231	0.4739	0.5261	0.9900
2^1	0.5432	0.4568	0.4957	0.5043	0.9972
2^0	0.5132	0.4867	0.5054	0.4946	0.9992
計					6.5230



(表に示す $p(0|0) \sim p(1|1)$ は上のシャノン線図に示す遷移確率を表している。)

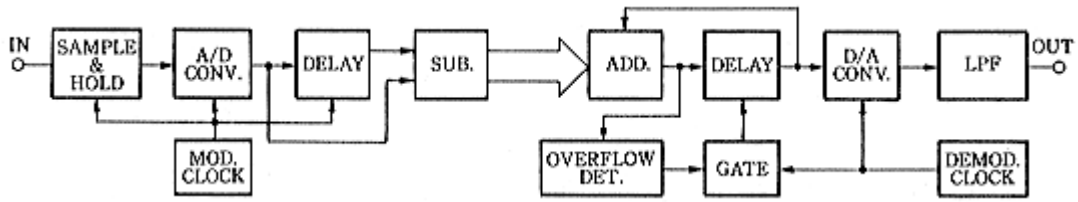


図4 差分PCM

広帯域音響信号の線形予測についてはここでは詳しくふれないが、予測値として1標本前の標本値をそのまま使う、すなわち $a_1 = -1$, 他の予測係数をすべて零とした単純な差分PCMでも通常の楽音に対し、16bitで量子化された信号が1標本当たり4~5bit, 本格的な予測符号化では10bit以上の信号を何ら劣化させない節減(伝送路と信号の整合)が可能である。

図4に示す予測値を1標本前の標本値とした線形予測を1階差分PCM, その予測誤差にさらに1階差分PCM線形予測を施す処理を2階差分PCMと呼ぶことにする。

差分PCMは微分操作と等価であるから、伝送可能な最大振幅は図5

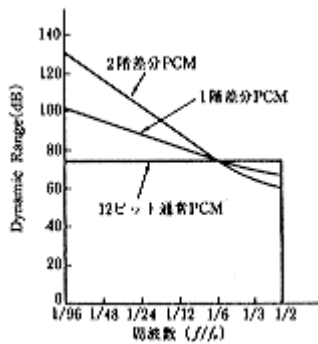


図5 差分PCMのダイナミックレンジ

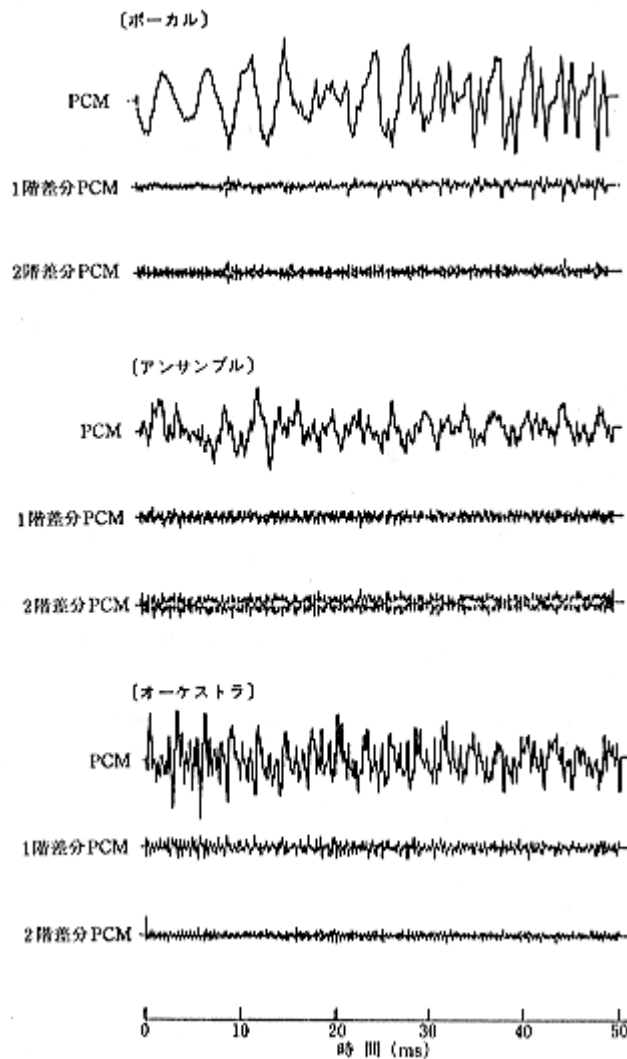


図7 差分PCMによる波形の変化

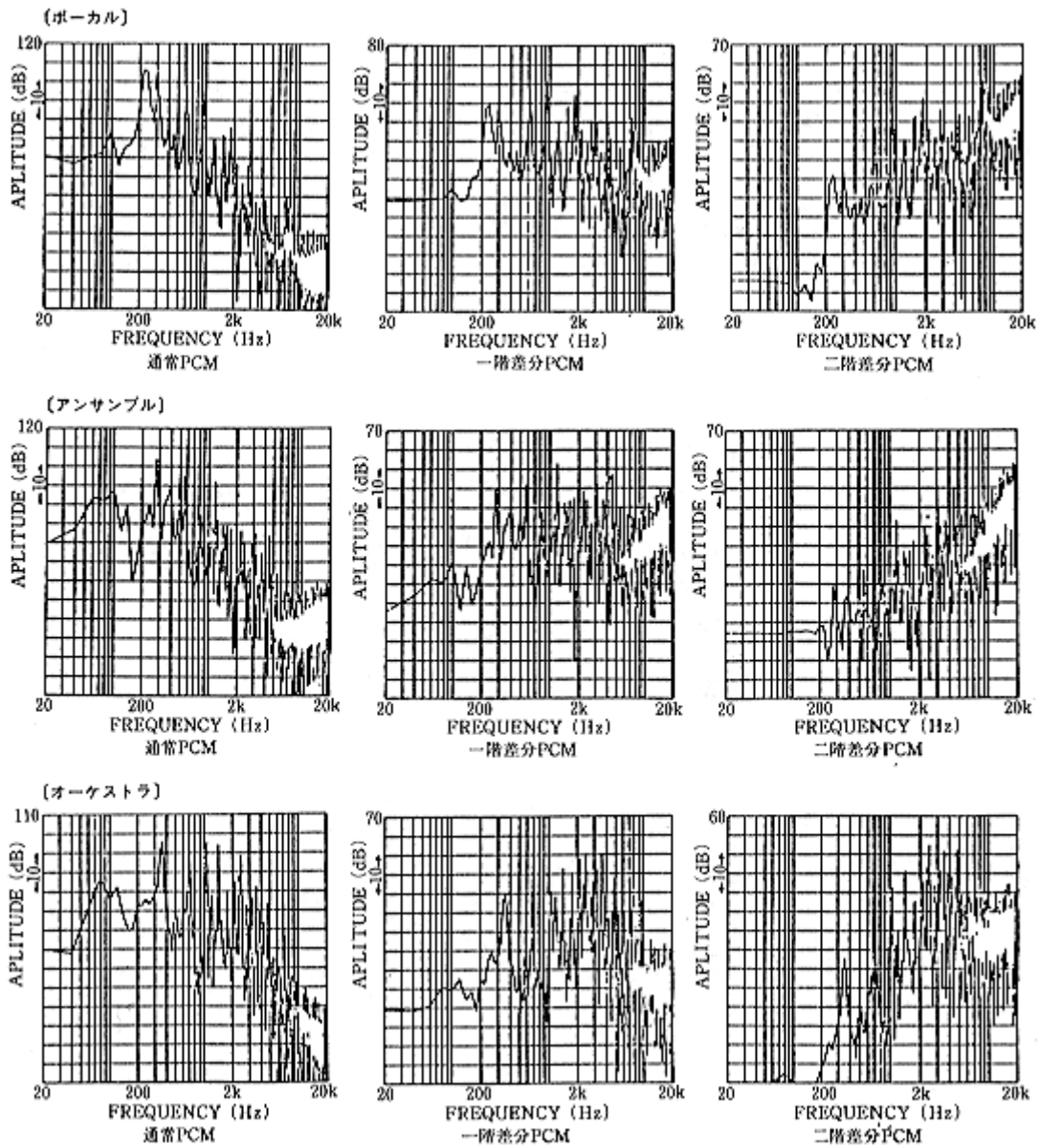


図8 差分PCMによるスペクトルの変化

エントロピー符号化

ところで、時間関数が除かれた予測誤差信号自体のレベル分布は図9に示すようにガウス分布に近い形をしている。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、伝送ビットレートを平均情報量の与える限界 (Rate Distortion Boundary) にさらに近づくことが可能である。エ

ントロピー符号化としては、シャノンの符号化やハフマン符号化があるが、ここでは比較的復号の容易なハフマン符号化をとりあげた。

表5に前述の2階差分PCMのレベル分布に基づき5bit、32ステップ

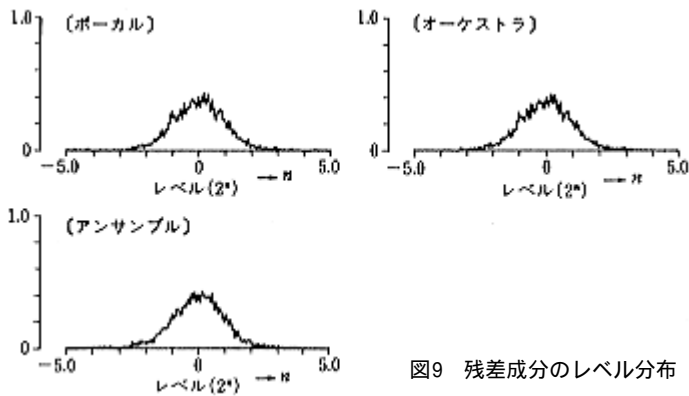


図9 残差成分のレベル分布

を対数近似したハフマン符号を示す。この結果1標本当たりの平均使用ビット数は2.979bitとなりレートディストーション限界の2.94bitにはほぼ一致している。ただしエントロピー符号は不等長符号であるため、伝送に使用する場合は大容量の一時記憶素子と辞書（符号割り当て表）が必要である。

また予測符号化やエントロピー符号化では複数の標本から信号が再現されるので、伝送路に誤りがあるとその影響が多くの標本値に及ぶので、システム的设计に当たってはこの点にも留意しなければならない。

浮動小数点方式の導入

信号の冗長度を利用して、無ひずみで伝送路の高度利用を行う方法について述べたが、次に人間の聴覚を利用した伝送路の節約について考えてみよう。

楽音や音声信号のレベルは、時間とともに大幅に変化するのは良く知られている。この性質を利用して標本ごとの振幅に応じて量子化ステップを変化させる浮動小数点方式は瞬

時圧伸とも呼ばれている。これをAD変換に際して行うと非一様量子化となる。さらに、広帯域音響信号のレベルの変動速度が比較的ゆるやかな点に注目して、信号を5～100

msの区間に分割しレベルの小さい区間は量子化ステップを小さく、大きな区間は量子化ステップを大きくとるブロック浮動小数点方式は準瞬時圧伸と呼ばれる。図10に示すように浮動小数点方式では各標本ごとに仮数情報 M [bit] と指数情報（レベル情報） L [bit] を同時に、すなわち1標本当たり $L+M$ [bit] 伝送する必要がある。これに対しブロック浮動小数点方式では、標本値列を N 標本ずつのブロックに分割し、それぞれのブロック内の標本の最大値からそのブロック内の量子化ステップを決定し、このステップでブロック内すべての標本の量子化を行うので、仮数情報 M [bit]、指数情報 L [bit] とすると、1標本当たりの平均符号長は浮動小数点方式の $L+M$

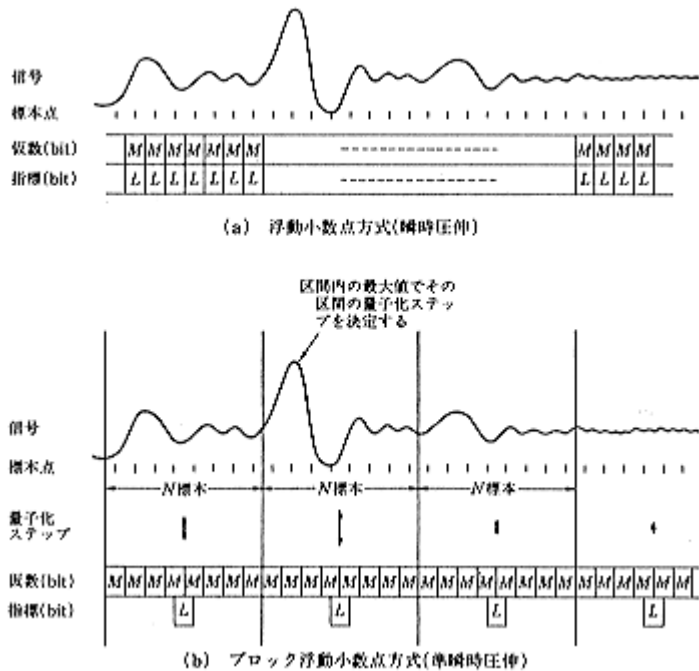


図10 浮動小数点方式とブロック浮動小数点方式の語構成

表5 各レベルの出現確率とハフマン符号

レベル n (dB)	出現確率 $p(n)$	情報量 $p \log_2 p$	ハフマン符号	ビット数 M_n	$M_n \times p(n)$
1	0.4290	0.5238	0	1	0.4290
2	0.1716	0.4364	110	3	0.5148
3	0.0977	0.3278	100	3	0.2930
4	0.0644	0.2547	1011	4	0.2574
5	0.0460	0.2042	11110	5	0.2298
6	0.0345	0.1677	10101	5	0.1727
7	0.0269	0.1402	111111	6	0.1613
8	0.0215	0.1189	111011	6	0.1287
9	0.0175	0.1019	111001	6	0.1047
10	0.0144	0.0881	101000	6	0.0864
11	0.0120	0.0767	1111100	7	0.0841
12	0.0101	0.0671	1110100	7	0.0708
13	0.0086	0.0589	1110001	7	0.0600
14	0.0073	0.0519	1010011	7	0.0512
15	0.0063	0.0458	11111011	8	0.0500
16	0.0054	0.0405	11101011	8	0.0429
17	0.0046	0.0358	11100001	8	0.0368
18	0.0040	0.0316	11100001	8	0.0317
19	0.0034	0.0279	10100100	8	0.0272
20	0.0029	0.0245	111110100	9	0.0262
21	0.0025	0.0215	111010100	9	0.0223
22	0.0021	0.0187	101001011	9	0.0190
23	0.0018	0.0162	1111101011	10	0.0178
24	0.0015	0.0139	1111101010	10	0.0148
25	0.0012	0.0118	1110101010	10	0.0122
26	0.0010	0.0099	1010010101	10	0.0099
27	0.0008	0.0080	11101010111	11	0.0086
28	0.0006	0.0063	11101010110	11	0.0065
29	0.0004	0.0047	10100101001	11	0.0046
30	0.0003	0.0032	101001010001	12	0.0032
31	0.0001	0.0016	101001010000	12	0.0015
32	0.0	0.0	————	0	0.0
	$\sum p(n)$ = 1.000	$\sum p \log_2 p$ = 2.9400 (bit)			$\sum (M_n \times p(n))$ = 2.9790 (bit)

に対し $(L/N) + M$ [bit/標本] となる。 N を大きくとれば大幅なビット節減が可能であるが、その区間のレベルの小さい部分での量子化雑音の増加は不可避である。

量子化雑音は量子化ステップあるいはその整数倍に一様分布する確率変数をディザとして重畳したうえ量子化し、量子化出力からこれを引き去ることにより、量子化ステップ幅に一様分布する白色性の雑音にすることが可能である。しかるに、非一様量子化における量子化雑音はたとえ最適なディザが重畳、減算されていても量子化ステップが変化するので、信号のレベルに応じて変化してしまい、入力信号との相関は不可避である。従ってこの変化が聴感上障害とならないようにパラメータを設定しなくてはならない。

浮動小数点方式、ブロック浮動小数点方式ともに原信号にある程度のひずみを与えるものであるが、人間の聴覚上の影響のない範囲でのビット節減は現実的には十分意味をもつ。表6に、各手法によりパラメータを変化させた際の S/N_q を示す。ここで s としては実験を行った際に用いた 12bit 一様量子化 PCM 伝送系により得られた信号の実効値

表6 ブロック不動小数点方式の S/N_q

仮数 (ビット)	瞬時圧伸制御時 の S/N (dB)	準瞬時圧伸制御時の S/N (dB)			
		$N=16$ 0.39 (ms)	$N=256$ 6.25 (ms)	$N=512$ 12.5 (ms)	$N=1024$ 25 (ms)
1	8.42	4.88	2.18	1.73	1.41
2	13.6	10.5	7.10	6.35	5.72
3	19.4	16.5	12.8	12.0	11.3
4	25.4	22.6	18.7	17.8	17.1
5	31.2	28.8	24.8	23.8	23.0
6	38.0	35.0	31.0	30.0	29.2
7	45.0	41.8	39.4	36.4	35.6
8	52.9	49.4	44.4	43.4	42.4

表7 ROMの容量と記録時間の関係

サンプル値データ ROM容量 (kワード)	指数データROM容量(ワード)				音声出力時間(s)		
	ブロック長N				サンプリング周波数f _s (kHz)		
	32	64	128	256	6	8	10
16	512	256	128	64	2.73	2.05	1.64
32	1024	512	256	128	5.46	4.10	3.28
48	1536	768	384	192	8.19	6.14	4.92
64	2048	1024	512	256	10.92	8.19	6.55

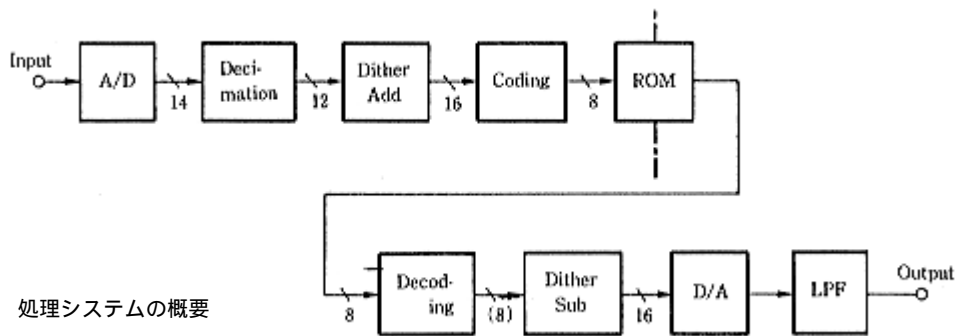


図11 処理システムの概要

を、 N_q としては各処理を施すことにより生じた信号と元信号との誤差の実効値を用いた。

40.96kHzで標本化、12bitで一樣量子化された信号として比較して品質劣化が認められない限界を各種の楽音について求めると、浮動小数点方式では仮数情報6bit、レベル情報3bit、すなわち9bit/標本である。ブロック浮動小数点方式では $M=6, L=3, N=256$ で $3/256 + 6 = 6.012\text{bit/標本}$ と大幅な節減が可能であった。 $N=256$ ということは6.25msごとにレベル検出を行うということで、これは、人間の耳の時定数と対応して考えても妥当な値といえよう。

一方、女性アナウンスについては、標本化周波数を8kHzとしたとき、 $M=4, L=2, N=64, (2/64) + 4 = 4.03\text{bit/標本}$ で8bit一樣量

子化とほぼ遜色ない音質が得られた。

予測符号化等とこの圧伸によるビット節減は、聴感上相違がなくとも、前者は、無ひずみであるのに、後者はあくまでも聴感上許容できるだけで、ひずみのあることを忘れてはならない。しかし、圧伸によるビット節減はかなり効果がある。音響信号に見合った量子化特性を見出すことの必要性を示唆しているともいえよう。

面体記憶素子への音響信号の記録

近年、線形予測符号の手法を用いた音声合成装置が集積回路化され、多方面への応用が検討されている。一方前述のように、比較的簡単な処理で符号化された音声波形データや楽音をROM等の半導体記憶素子

に記録しておき必要な都度読み出す、いわば固体録音再生方式も半導体記憶素子の集積度の向上と価格や消費電力の低減とともに実用化されつつある。

固体記憶素子への録音・再生、波形符号化方式の主な利点は次のとおりである。

- (1) 比較的良い自然な音質が得られるうえ、記録するデータ作成が簡単であり、ユーザレベルでのデータ書き込みも比較的容易である。
- (2) 音声信号の再生手順が単純であり、専用チップを用いなくても安価に音声出力回路を実現できる。
- (3) 一人の音声に対してだけでなく、複数の人の声や音楽あるいはそれらが混合された信号にも適用可能である。

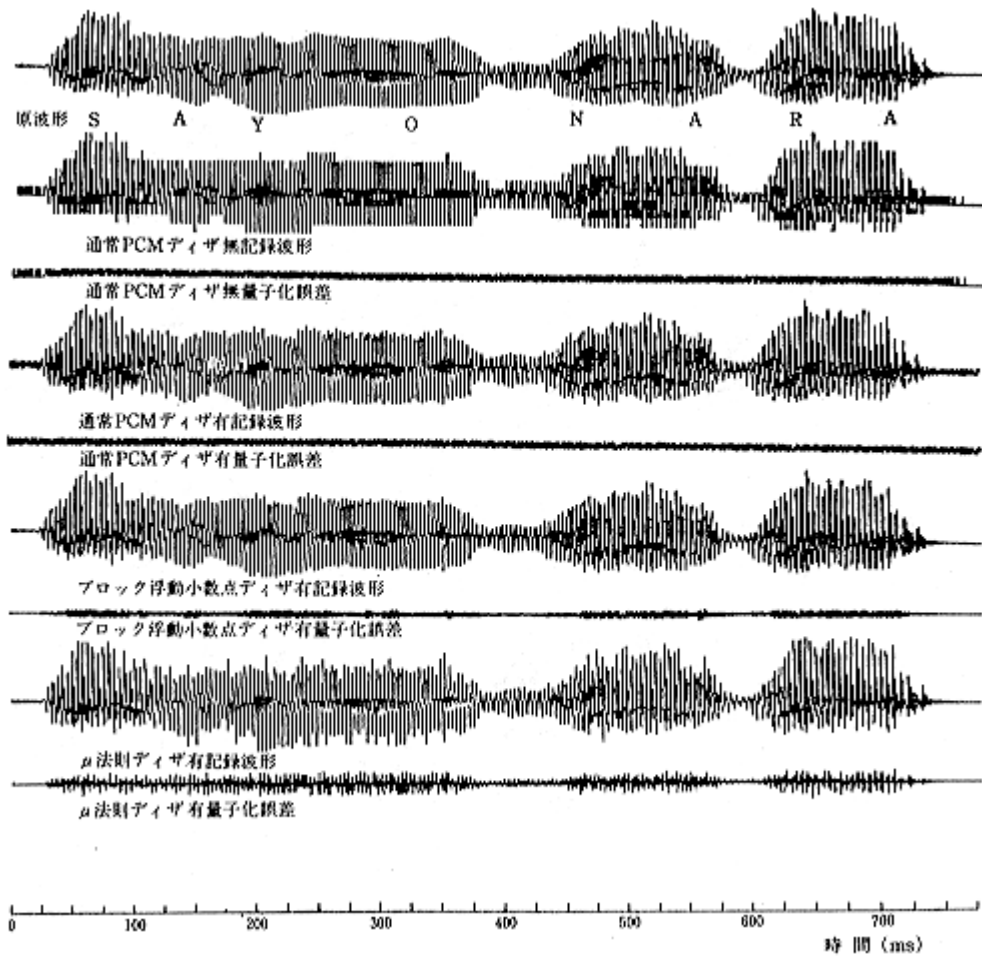


図12 処理結果

反対に欠点は、記録する信号の容量を線形予測による方法でのパラメータほどには低くできない点である。しかし、固体記憶素子の集積度の向上と価格の低減は今後も続くものと考えられるので、将来的には著しいデータ圧縮よりは上記の長所が重視されることとなる。

無人気象通報装置への応用

無人気象通報装置に用いる女性アナウンスをROMに記録する処理を行う機会を得た。これはテレメー

タからの情報を使って従来はアナログ録音された磁気ドラムのトラック選択することにより船舶に向けて無人の気象通報を行っていたシステムにおいて、可動部分を持つドラム装置を固体記憶素子を用いたデジタル記録に置き換えようという試みで、1981年から海上保安庁において順次採用されている。

各種の処理方法

スタジオにおいてPCM録音(標本化周波数 44.1kHz, 量子ビット数

14)した女性アナウンスを標本化周波数 1/5 (8.82kHz)で64kbit ROM (8bit × 8192)に1秒弱記録する手法について、ここでは特に復調時のハードウェア構成の簡単な下記的手法について検討を加えた。

(1) 通常PCM

一様量子化した音声信号をそのままROMに固定する方法。アナログ段でのプリエンファシス、ディエンファシスの併用は考慮する。

(2) 浮動小数点(瞬時圧伸)

信号の振幅により量子化ステップ

を2のべき乗で変化させる，いわゆる瞬時圧伸。サンプルごとに仮数と指標が必要。

(3) ブロック浮動少数点(準瞬時圧伸)

浮動少数点各サービスごとに量子化ステップを選択するのに対し，複数サンプルを1ブロックと考え，その中の最大値で量子化ステップを決定する，いわゆる準瞬時圧伸。データROMの一部に指標をまとめて書き込むことも可能である。

(4) μ 法則

μ 法則で圧伸則を決定する方法。データROMの一部に圧伸則の情報を記録しておくことが可能。さらに再生時には変換ROMテーブルにより伸長が可能。

(5) 準瞬時 μ 法則

(3)と(4)を組み合わせた方式。

女声アナウンス“さよなら”を上記の各種処理方式で処理した。処理の手順を図12に示す。元のアナウンスが14bitでPCM録音されているので，図12に示すように符号化に際し擬似ランダムディザをデジタル段で付加し，16bit DA変換器を用いることにより復調時にディジタ

表8 誤差電力

		通常PCM	浮動小数点	ブロック浮動小数点	μ 法則
$\overline{S/D}$ (dB)	ディザ有	40.35	37.79	42.64	38.39
	ディザ無	40.41	37.73	42.54	38.40
$\overline{S/D}$ (dB)	ディザ有	4.47	30.64	13.80	30.35
	ディザ無	2.04	30.77	8.17	31.22

S; Signal
 D; (Signal) - (Processed Signal)
 N; Number of Data
 $\overline{S/D}; 10\text{Log}(\sum S^2 / \sum D^2)$ (dB)
 $S/D; 10\text{Log} \sum (S/D)^2 / D$ (dB)
 ブロック浮動小数点: (ブロック長64)
 μ 法則: $\mu = 200$
 Normalized μ Function;
 for $-1 \leq x \leq +1$ $F(x) = \text{sgn}(x) \frac{\ln(1 + \mu|x|)}{\ln(1 + \mu)}$

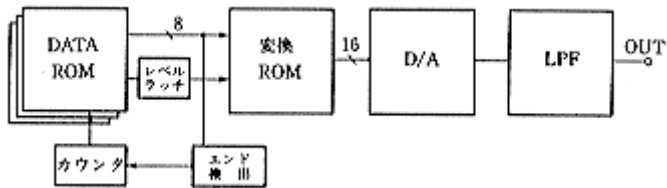


図13 再生用ハードウェアの構成

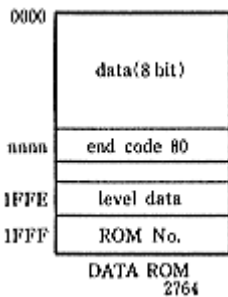


図14 データROMの内容

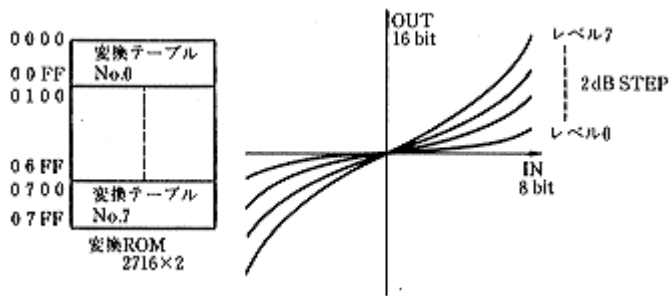
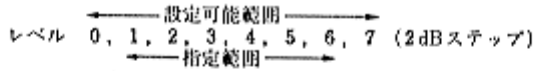


図15 変換ROMの内容と変換特性

表9 メッセージの内容と各々のレベル情報1



NO	メッセージ	レベル	備考
76	23時	1	
77	55分	2	
78	39分	4	
79	北の風	5	
80	北西の風	2	
81	西の風	3	
82	南西の風	5	

NO	メッセージ	レベル	備考
109	気象	3	
110	状況を	3	
111	お知らせ	3	
112	します	3	
113	繰り返します	4	
114	さようなら	3	
115	はこ	4	液高

ル段でこのディザを引き去る処理を採用した。

信号と量子化雑音の比の平均と、それぞれの長時間平均の比を表8

に、処理した波形と誤差波形を図13に示す。なおディザの有無についても併記した。

実験の結果、メッセージごとにレ

ベル情報を記録する、準瞬時 μ 法則が有利なことが明らかとなった。

ハードウェアの構成を図13に示す。

かなり高品質の1秒弱の音声を64kbitのROMに固定し、約100種類の通報の可能な小規模な無人気象通報装置が実現できた。

図14に音声データROMのデータ構成を、図15に変換ROMの内容と圧伸特性を示す。なお、ビット削減時直流誤差が蓄積されないように、LSBの処理を行っている。また、表9にメッセージの内容の一部とそのレベルを示す。