Proceedings of ISCAS 85

A New Low Bit-Rate Waveform Coding Using Extrapolation and Predictive Coding

Kenji Nakayama and Toshihiko Mizukami

Transmission Division, NEC Corporation
Kawasaki, 211 Japan

ABSTRACT

This paper describes a new waveform coding method using extrapolation and predictive coding techniques. The proposed method can provide 80% bit-rate reduction compared to linear PCM coding.

I. INTRODUCTION

Low bit-rate coding technique is one of key factors in speech communication and storage.[1] This paper presents a new low bit-rate waveform coding method in which oversampling, predictive coding and extrapolation techniques are effectively combined to reduce bit-rate.

Oversampling technique is used to convert a input signal into highly correlated signal. Predictive coding provides low bit-rate encoding scheme for such a oversampled signal.

In addition to these techniques, extrapolation is introduced in a decoder. If a signal is band limited, we can reconstruct an original signal from a part of the signal by extrapolating the rest. Hence, we need transmit only a part of signal samples, not a whole samples. Applying the extrapolation to the decoder input signal, the proposed coding method can reduce the bit-rate by 80% over conventional linear PCM coding.

II. CONFIGURATION OF CODEC

Fig.1 shows a block diagram of the proposed coder and decoder, and Fig.2 shows the signal waveforms in each stage.

A. CODER

Let x(n) be a band limited signal with band width $0 \leq f \leq f_b$, sampled at the sampling frequency $f_s = 2f_b$. By inserting M-1 zeros between each sample of x(n), we obtain u(m) with the sampling rate

$$f_{sh} = M \cdot f_s \qquad (1)$$

where M is the oversampling ratio of u(m) over x(n).

The over sampled signal u(m) is separated into section of length Nw by multiplying the rectangular window w(m). The i-th finite duration sequence $\hat{u}_i(m)$ is then filtered by a low pass filter (LPF), whose passband is $0 \leq f \leq f_b$. The LPF eliminates the unnecessary frequency components and retaines only the baseband signal within $0 \leq f \leq f_b$. The filter output $v_i(m)$ is an interpolated version of $\hat{u}_i(m)$ and has a sequence of length Nv:

$$Nv = Nw + Nh - 1 \qquad (2)$$

where Nh is LPF length. By extracting successive Ns(=Nv/M) samples of $v_i(m)$, we obtain $\hat{v}_i(m)$. Next, $\hat{v}_i(m)$ is coded by a predictive coder and produces the coder output y(m) of length Ns.

B. DECODER

In the decoder, input signal y(m) is converted to $\hat{v}_i^*(m)$ by a predictive decoder. $\hat{v}_i^*(m)$ is a replica of $\hat{v}_i(m)$ in the coder. Since $v_i(m)$ in the coder is band limited within $0 \leq f \leq f_b$, it is possible to reconstruct $v_i^*(m)$ of length Nv from $\hat{v}_i^*(m)$ of length Ns by extrapolation using their linear dependent relations based on the discrete Fourier transform (DFT). Though $v_i^*(m)$, reconstructed version of $v_i(m)$, is of length Nv, the final decoder output is needed only at every M-th sample, because the sampling rate at the decoder output is f_s. Hence, we need calculate $v_i^*(m)$ only at $m = M \cdot n$ to obtain $\tilde{v}_i^*(n)$. $\tilde{v}_i^*(n)$ is of length Ns, and has a overlapping portion with the next section $\tilde{v}_{i+1}^*(n)$. Final decoder output x(n) is obtained by overlapping-addition of the adjacent $\tilde{v}_i^*(n)$'s.

III. DESIGN CONSIDERATION

This section discusses the design of each functional block in detail.

A. LOW PASS FILTER

An LPF is used to band limit the zero interpolated signal u(m). In order to extrapolate $v_i^*(m)$ in the decoder using the dependent relations of the DFT, the filter output $v_i(m)$ must be a finite duration sequence. Hence, the LPF must be a finite impulse response (FIR) filter with passband 0 - f_b Hz and a sampling frequency $M \cdot f_s$ Hz.

Required filter specifications for speech signal are:

passband : 0 - 3400 Hz
stopband : 4000 - 8000 x M Hz
passband ripple Ap : as low as possible
stopband attenuation As : as high as possible

The filter length Nh is roughly proportional to M for given Ap and As.[2] For Ap = ±0.05dB and As = 80dB, Nh is given by

$$Nh \simeq 45M \qquad (3)$$

Fig.3 illustrates a block diagram of the LPF. Since $\hat{u}_i(m)$ has non-zero values only every M-th samples, we need only Nh/M multiplications per output sample. Moreover, the LPF output $v_i(m)$ is required by only Ns samples, we can save computation by Ns/Nv.

CH2114-7/85/0000-0317$01.00 © 1985 IEEE

B. PREDICTIVE CODER/DECODER

The transfer function Hp(z) of the coder is chosen as

$$Hp(z) = (1 - z^{-1})^m, \quad 0 \leqq m \leqq K-1$$
$$= (1 - z^{-1})^K, \quad K \leqq m \leqq Ns-1 \qquad (4)$$

where m is the time index of $\hat{v}_i(m)$ and K is a predictor order. This Hp(z) gives a robust predictor in the sence that predictor coefficients are independent of the input signal statistics and the sampling frequency[3]. The predictor order K is one of important design parameters. To evaluate the order K, let us consider the case where coder input is sinusoidal signal of frequency f_b, which is the most severe frequency among all of in-band frequency components.

Let the coder input signal be

$$\hat{v}_i(m) = \cos(m\omega_b + \phi) \qquad (5)$$

where $\omega_b = 2\pi f_b/Mf_s = \pi/M$.

Then, the output y(m) of the coder Hp(z) is

$$y(m) = A^{(m)}\cos(m\omega_b + \phi^{(m)}), \quad 0 \leqq m \leqq K-1$$
$$= A^{(K)}\cos(m\omega_b + \phi^{(m)}), \quad K \leqq m \leqq Ns-1 \qquad (6)$$

where $A^{(m)} = (2\sin(\pi/2M))^m$ and $\phi^{(m)}$ is constant, function of m and ω_b. Hence, the magnitude of y(m) decreases as K is increased. Let w_i be a word length of the input $\hat{v}_i(m)$ with sign bit, then the required word length $w_o(m)$ for the output of m-th order predictive coder including sign bit is

$$2^{(w_o(m)-1)} = (2\sin(\pi/2M))^m / 2^{-(w_i-1)}$$

So

$$w_o(m) = m\log_2(2\sin(\pi/2M)) + w_i \qquad (7)$$

K is chosen so that $w_o(K)$ in (7) is equal to or less than 1.

The total number of bits to represent $\hat{v}_i(m)$ of length Ns is given by

$$N_T = \sum_{m=0}^{K-1} w_o(m) + Ns - K$$
$$= Kw_i + \frac{K(K-1)}{2}\log_2(2\sin(\pi/2M)) + (Ns-K) \qquad (8)$$

Circuit configuration of the predictive coder and decoder are shown in Fig.4. In this figure, a predictor Pm(z) is given by

$$Pm(z) = 1 - Hp(z)$$
$$= -\sum_{k=1}^{m}(-1)^k\binom{m}{k}z^{-k}, \quad 0 \leqq m \leqq K-1$$
$$= -\sum_{k=1}^{K}(-1)^k\binom{K}{k}z^{-k}, \quad K \leqq m \leqq Ns-1 \qquad (9)$$

where $\binom{m}{k}$ is the binomial coefficient.

C. EXTRAPOLATOR

Band limited signal can be reconstructed from a part of original signal samples using an extrapolation technique[4]. Let the DFT of $v_i(m)$ be $V_i(\ell)$. Since the $v_i(m)$ is band limited within $0 \leq f \leq f_b$ by the LPF,

$$V_i(\ell) = \sum_{m=0}^{Nv-1} v_i(m)\exp(-j2\pi\ell m/Nv) = 0, \qquad \pi/M \leqq 2\pi\ell/Nv \leqq \pi \qquad (10)$$

Rewriting real and imaginary components separately,

$$\sum_{m=0}^{Nv-1} v_i(m)\cos(2\pi\ell m/Nv) = 0$$
$$\sum_{m=0}^{Nv-1} v_i(m)\sin(2\pi\ell m/Nv) = 0 \qquad \pi/M \leqq 2\pi\ell/Nv \leqq \pi \qquad (11)$$

Eqn.(11) gives (M-1)Nv/M linear equations with Nv variables. Hence, if Nv/M elements of $v_i(m)$'s are known, we can solve the equations about the remaining (M-1)Nv/M $v_i(m)$'s. By a matrix representation, the unknown $v_i(m)$'s are given by

$$\underline{v}_u = \underline{c}\cdot\underline{v}_k \qquad (12)$$

where $\underline{v}_u$: column (M-1)Nv/M - vector of unknown $v_i(m)$'s,
$\underline{v}_k$: column Nv/M - vector of known $v_i(m)$'s,
$\underline{c}$: (M-1)Nv/M x Nv/M coefficients matrix.

Each entry of the matrix $\underline{c}$ can be calculated from the DFT coefficients $\exp(-j2\pi\ell m/Nv)$ if the time index m of known $v_i(m)$'s are specified.
Since the final decoder output is needed only at every M-th point, only $v_i^*(m), m = n\cdot M, 0 \leq n \leq Ns-1$, is sufficient to be computed in the extrapolator.

IV. BIT-RATE COMPRESSION RATIO

The bit-rate compression ratio ρ is defined as the ratio of bit-rate at the coder output over the input signal bit-rate. Let the word length of the input x(n) be w_i, then the bit-rate of input signal is

$$B(\text{input}) = w_i \cdot f_s \text{ bits/sec.} \qquad (13)$$

From the previous discussion, the coder output requires N_T bits to transmit y(m) in every Nw/f_{sh} second, then the bit-rate of the output is

$$B(\text{output}) = N_T/(Nw/f_{sh}) = Mf_sN_T/Nw \text{ bits/sec.} \qquad (14)$$

Assuming $v_i(m)$ in the coder has the same word length as that of input x(n), the compression ratio ρ is given by

$$\rho = \frac{B(\text{output})}{B(\text{input})} = MN_T/w_iNw \qquad (15)$$

Numerical examples for ρ are shown in Fig.5 and parameter values are tabulated in Table 1 and 2. As shown in Fig.5, the proposed method can reduce the bit-rate to 1/5 - 1/8 compared to linear PCM coding. Required number of multiplications per second O(MPL) is given by

$$O(MPL) = f_{sh}(NhNs/MNw + 2NsK/Nw + Ns(Nv-Ns)/MNw)$$
$$\simeq f_s Nv(Nh/M + 2K + Nv/M)/Nw \qquad (16)$$

O(MPL) for the example in Fig.5 a) is around 280 x f_s for each case. This number is acceptable for actual hardware implementation.

V. NOISE EVALUATION

Noise sources in the proposed method are:
(1) Finite stopband attenuation of LPF
(2) Quantization error in the predictive coder
(3) Extrapolation error
(4) Round-off error in the LPF computation.

(4) can be neglected if the signal word length in LPF is appropriately chosen.

Signal components in the stopband of the LPF fall into baseband by M:1 decimation. This noise can be expressed as

$$N_1 = (M-1)10^{-As/10}\frac{1}{2\pi}\int_{-\pi/M}^{\pi/M}|U(e^{j\omega})|^2 d\omega \qquad (17)$$

Next, quantization error in the coder is

$$N_2 = 2^{-2(w_i-1)}/12 \qquad (18)$$

where w_i is the word length of the coder input.

In eqn. (11), infinite loss of As is assumed. However, this assumption is not actually satisfied and the extrapolation causes error to $v_i^*(m)$. Let the extrapolation error be r(m), then

$$\sum_{m=0}^{Nv-1} r(m)\exp(-j2\pi\ell m/Nv) = V(\ell) = H(\ell)\,U(\ell) \qquad (19)$$

where $H(\ell)$ is an LPF transfer function and $U(\ell)$ is the DFT of u(m).

By taking squared-summation over ℓ from Nv/2M to Nv-Nv/2M-1,

$$\frac{M-1}{M}Nv\sum_{m=0}^{Nv-1} r^2(m) = 10^{-As/10}\sum_{\ell=Nv/2M}^{Nv-Nv/2M-1}|U(\ell)|^2 \qquad (20)$$

Hence, the extrapolation error is given by

$$N_3 = \sum_{m=0}^{Nv-1} r^2(m) \simeq M10^{-As/10}\frac{1}{2\pi}\int_{-\pi/M}^{\pi/M}|U(e^{j\omega})|^2 d\omega \qquad (21)$$

Total noise power N is

$$N = N_1 + N_2 + N_3 = (2M-1)10^{-As/10}\frac{1}{2\pi}\int_{-\pi/M}^{\pi/M}|U(e^{j\omega})|^2 d\omega + 2^{-2(w_i-1)}/12 \qquad (22)$$

Suppose the coder input is a sine-wave with amplitude A. The signal to noise ratio (SNR) for the input is given by

$$SNR = 10\log\left(\frac{A^2}{2}\Big/\left(\frac{A^2}{2}(2M-1)10^{-As/10} + 2^{-2(w_i-1)}/12\right)\right) \qquad (23)$$

Fig.6 shows the SNR characteristics for As = 70dB and w_i= 13bits. As shown in Fig.6, SNR is slightly degraded as M is increased, because the noise caused by finite stopband attenuation of the LPF becomes predominant for high level input.

VI. CONCLUSION

A new approach for low bit-rate coding is proposed. Numerical evaluation for the bit-rate compression ratio and SNR shows the proposed method provides good performance compared with the conventional linear PCM coding. Furthermore, since the predictor in the coder does not utilize any statistics of input signal, this method can be applied for both speech and music coding.

REFERENCES

(1) L.R.Rabiner and R.W.Schafer, Digital Processing of speech Signals, Prentice-Hall, Inc., 1978.
(2) L.R.Rabiner and B.Gold, Theory and Application of Digital Signal Processing, Prentice-Hall, Inc., 1975
(3) S.K.Tewksbury and R.W.Hallock, "Oversampled, Linear Predictive and Noise Shaping Coder of order 1," IEEE Trans. CAS-25, no.7, pp436-447, July 1978.
(4) T.Iijima, "Decision Method of Expansion Coefficients for Finite Fourier Series by Partial Sampling Values (in Japanese)," IECE of Japan, Trans. J64-A, no.3, pp219-226, March 1981.

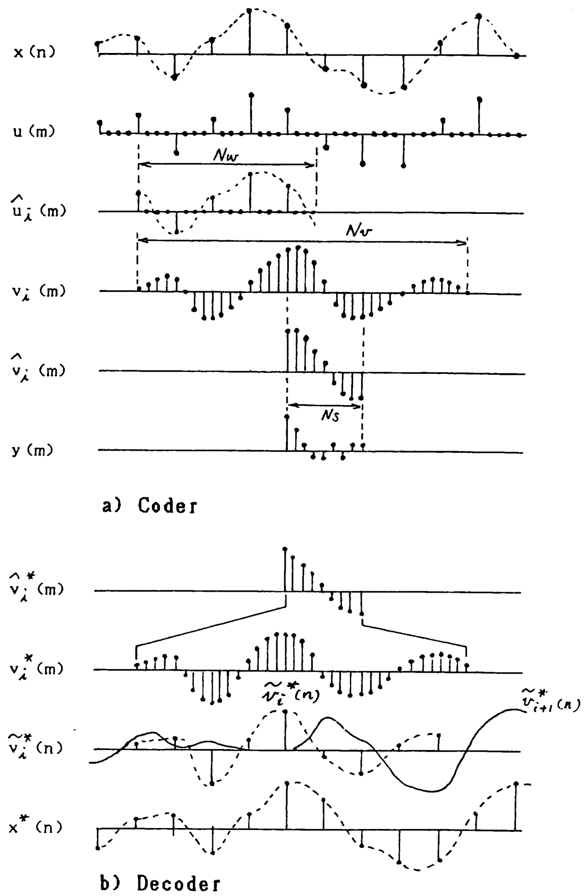


Fig.2. Signal waveforms.

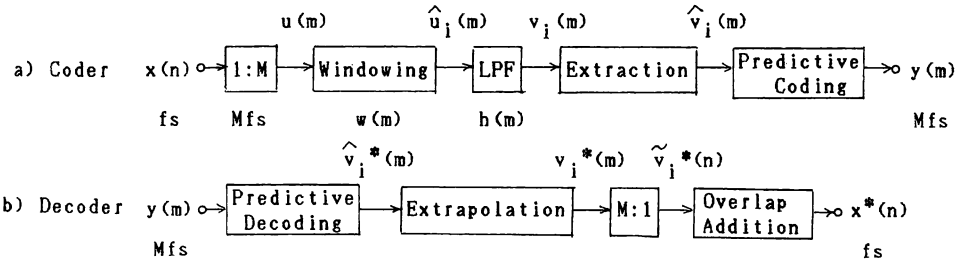
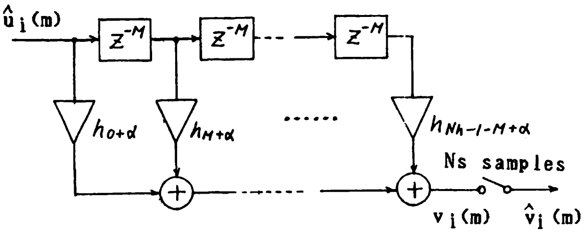
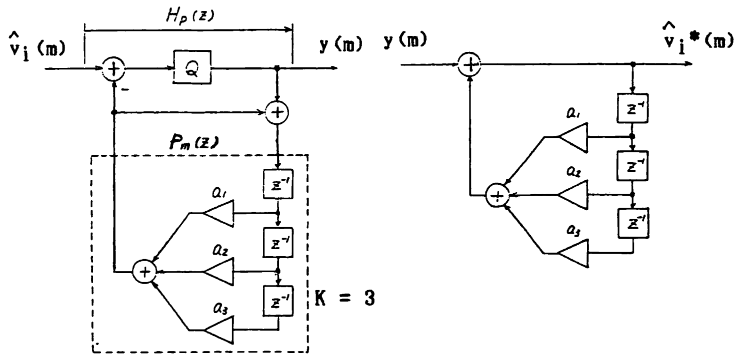


Fig.1. Block diagram of coder/decoder.



$\alpha = ((m))_M$: m modulo M

Fig. 3. LPF block diagram.



a) Coder b) Decoder

Fig. 4. Predictive coder/decoder.

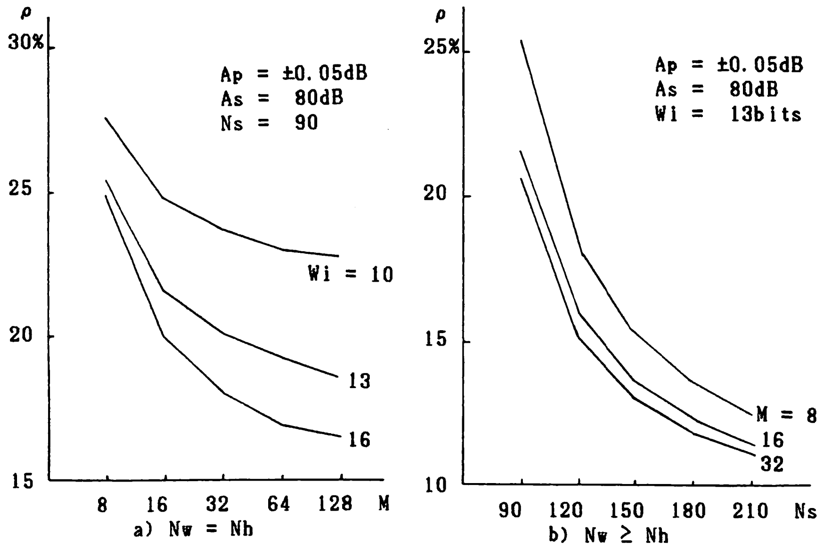


Fig. 5. Bit-rate compression ratio.

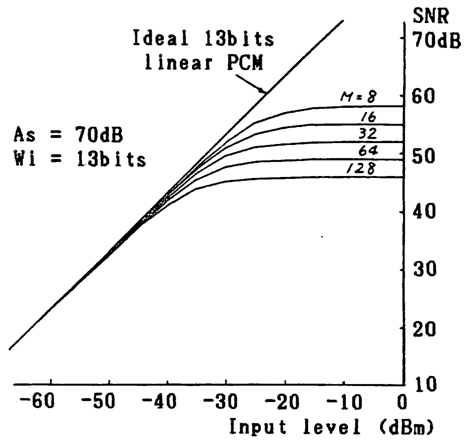


Fig. 6. SNR characteristics.

Table 1. Parameter values for example in Fig. 5 a).

M	Wi	K	Nw = Nh	Nv	Ns
8	10	7	360	719	90
	13	9			
	16	11			
16	10	4	720	1439	90
	13	5			
	16	6			
32	10	3	1440	2879	90
	13	4			
	16	4			
64	10	2	2880	5759	90
	13	3			
	16	3			
128	10	2	5760	11519	90
	13	2			
	16	3			

Table 2. Parameter values for example in Fig. 5 b).

M	Wi	K	Nh	Nw	Nv	Ns
8	13	9	360	360	719	90
				600	959	120
				840	1199	150
				1080	1439	180
				1320	1679	210
16	13	5	720	720	1439	90
				1200	1919	120
				1680	2399	150
				2160	2879	180
				2640	3359	210
32	13	4	1440	1440	2879	90
				2400	3839	120
				3360	4799	150
				4320	5759	180
				5280	6719	210