RECURSIVE AND ADAPTIVE PREDICTIVE CODING OF SPEECH

Kazuo NAKATA and Kin-ich HIGURE

Department of Electronics, Chiba Institute of Technology
(2-17-1, Tudanuma, Narashino-shi, Chiba, 275, JAPAN)

ABSTRACT

A new algorithm of speech coding " recursive and adaptive prediction "is proposed and tested. An adaptative linear prediction of input is carried out at sample by sample, and only predictive residuals are quantized and transmitted in binary codes Predictive coefficients are adaptively controlled by qauntized prediction error. Segmental SNR of almost 22 dB is obtained at 16 k b/s by the cascade connection of 2 stages of prediction. The algorithm can handel mixed voices as well, and easy be implemented by single DSP.

1. INTRODUCTION AND THE OBJECTIVES OF THE RESEARCH

Needs for an efficient coding of speech increase rapidly due to wide use of handy personal telphone and new multimedia communications. The requierements for speech coding in these new fields are [1,2] ;

1) good voice quality in less bit rate,
2) no time delay for coding,
3) simple in algorithm, low cost implementation by single DSP,
4) able to handle mixed voices , and
5) robust for noisy speech coding.

The purpose of the research is to develop coding method to meet these requirements. The basic algorithm and preliminary experiments are described in the section two. Some trials to improve its SNR are in the section three, and in the section four the results of mixed voices and noisy speech coding, and final conclusions are given in the section five.

2. THE BASIC ALGORITHM

The algorithm taken to meet the above requirements is recursive and adaptive predictive coding of speech.
Speech wave forms are assumed to be generated by AR process and being predictable in the following ways:

$$\hat{y}(n)=\Sigma\,\alpha(n,i)y(n-i): \quad \text{prediction.} \quad (1)$$

where $y(n)$: input, n : sample number,
p : the order of prediction,
$\{\alpha(n,i)\}$: $i=1,2,\ldots,p$ predictive coefficients

$$e(n) = y(n) - \hat{y}(n) \; : \text{prediction error or residual.} \quad (2)$$

If error $e(n)$ is given, predictive coefficients $\{\alpha(n,i)\}$ are adapted in following way to minimize square errors. [3]

$$\alpha(i,n+1) = \alpha(i,n) + \mu * y(n-i) / P(n): \quad \text{adaptation} \quad (3)$$

μ : adaptive gain,
$P(n) = \Sigma\; y(n-i)^2$

Then, the $e(n)$ is known both sending and receiving end, and $\{y(n-i)\}$ are replaced by the outputs : $\{Y(N-i)\}$, the input $y(n)$ is reproduced as an output:

$$Y(n) = \hat{y}(n) + \bar{e}(n) \; : \; \text{output} \quad (4)$$

$\bar{e}(n)$ is transmitted value of $e(n)$ after quantization $Q[e(n)] = \bar{e}(n)$.

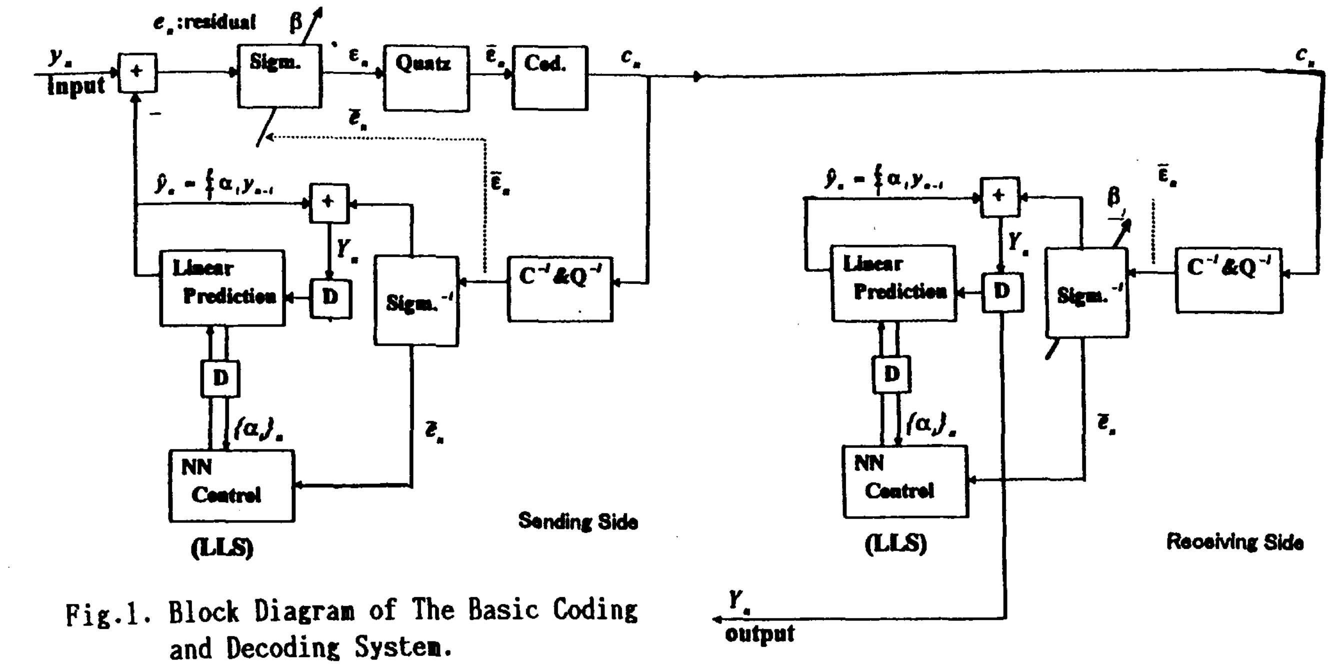

Fig.1. Block Diagram of The Basic Coding and Decoding System.

An adaptation of β in Sigmod transformation is used as a nonlinear quantization of error. The process is repeated sample by sample. The process is shwon in Fig.1. The series of experiments are carried out to confirm feasibilities of the above basic algorithm , and to determine the best value of adaptive gain μ and the order of prediction p.
The results of experiments show the best value of μ is 0.05 and p is 12, and the way of control of β is almost the same as the case of step control of ADPCM.
The segmental SNR is about 20 dB by 2 bit /sample or 16 k b/s of 8 k Hz sampling. They also show high segmental SNRs as 30 to 36 dB at 24 k b/s to 32 k b/s.

3. IMPROVEMENTS OF THE BASIC ALGORITHM

The segmental SNR 20 dB at 16 k b/s is good but not high enough compair to ADPCM Several variations of the basic one have been tried. The most effective one is the cascade connction of two stages of prediction of 6 order with different gains. The SNR reaches almost 22 dB. The block diagram of the system is shown in Fig.2. The SNR of some typical voices in various conditions are shown in the Table 1.

Table 1. Segmental SNRs in dB of Typical Voices for Various Conditions.

		2 bit/samp	3 bit/samp	4 bit/samp	5 bit/samp
1-stage	WORD (male)	19.77	26.47	32.97	38.30
	WORD (female)	19.09	25.54	31.10	36.62
p=12	SENTENCE (male)	19.63	27.21	33.20	38.66
2-stages	WORD (male)	20.26	27.73	34.46	40.10
	WORD (female)	19.91	26.93	33.46	38.87
p=6 *2	SENTENCE (male)	21.45	29.79	36.36	41.90

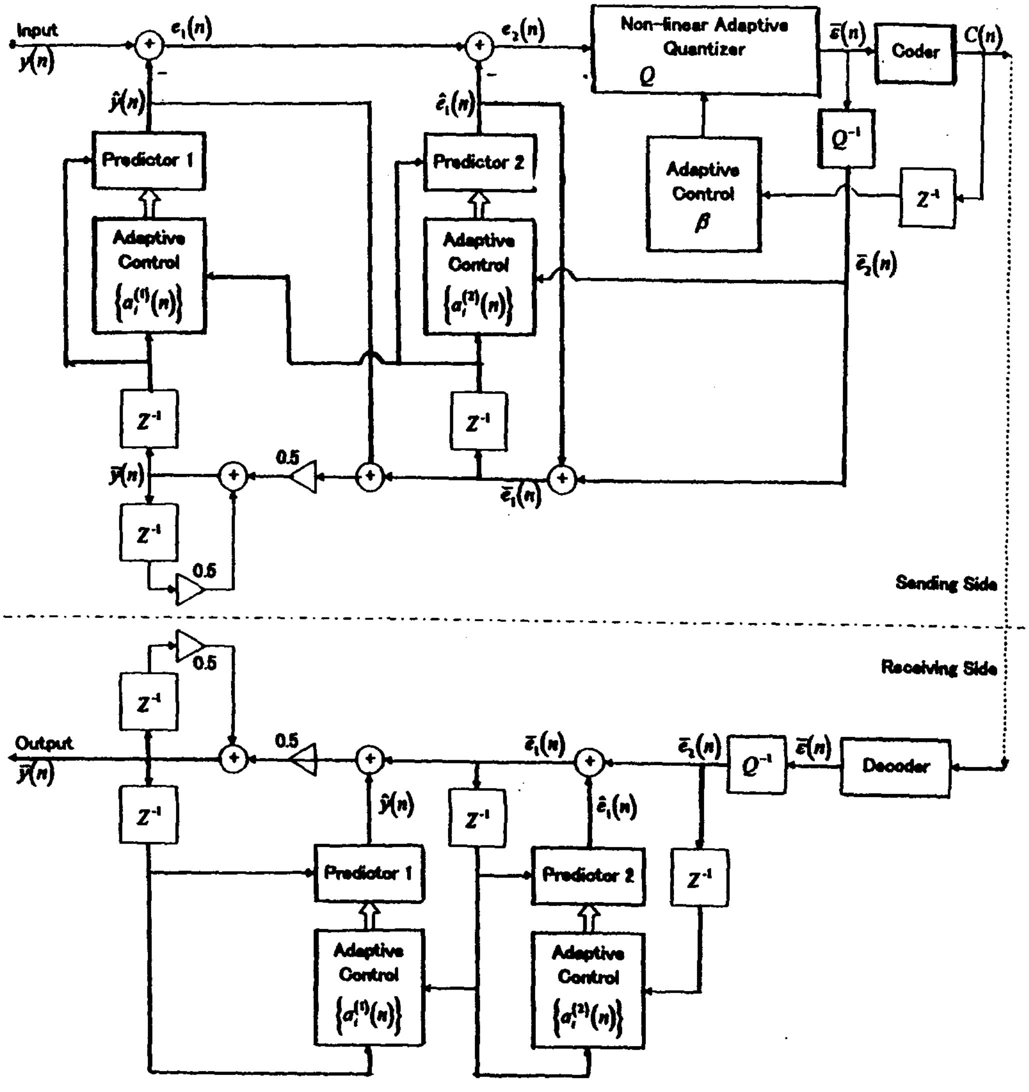

Fig.2. Block Diagram of The Improved Two-Stage Coding and Decoding System.

Table 2. Segmental SNRs in dB for Mixed Voices

		2 bit/sampl	Mixed	3 bit/samp	Mixed
1-st	Speaker A (male)	19.63		27.21	
	Speaker B (male)	19.63	20.30	27.21	27.81
2-st	Speaker A (male)	21.45		29.70	
	Speaker B (male)	21.45	21.60	29.70	30.16
1-st	Speaker A (male)	19.63		27.21	
	Speaker B (female)	15.24	17.48	22.38	24.82
2-st	Speaker A (male)	21.45		29.06	
	Speaker B (female)	15.80	18.34	23.77	26.00

A voice quality of the output of 16 k b/s is rather harsh, and improved aurally by the smoothing of outputs by averaging :

$\overline{Y}(n) = [\ Y(n) + \overline{Y}(n-1)]/2$: smoothing (5)

This improvement is confirmed by the decr eas of spectrum distance from input. For example, Cepstrum distances can be reduced from 5 dB to 4.3 dB by the averaging. This is no more true for the case of 24 k b/s, because SNR is high enough and harsh ness is not audible in themselves.

4 MIXED AND NOISY VOICES CODING.

In real situations, speaker A's voice co exists often with speaker B's voice, or voices comes to the microphone with some background noise, and the algorithm must handel these case properly.
The series of experiments are carried out to test the case. The Table 2 shows the results, the algorithm can properly handel the case. For mixed voice, SNRs do not degrade compaired to the SNR of one voice alone. For noisy input voice, SNRs can not be improved, but can keep advantages to ADPCM.

Fig. 1. Block Diagram of The Basic Coding and Decoding System.

Fig. 2. Block Diagram of The Improved Two-Stage Coding and Decoding System.

Table 1. Segmental SNRS in dB of Typical Voices for Various Conditions.

Table 2. Segmental SNR in dB for Mixed Voices.

5. CONCLUSIONS

The new algorithm of "recursive and adapt ive predictive" coding of speech is proposed and tested. By the cascade connect ion of two stages of prediction of order 6 with differnt adaptive gains, SNR better than 21 dB is obtained.
The algorithm is so simple that easy to implement by a single DSP, and needs no time delay for coding.
The algorithm is also applicable to mixed voices coding without any increase of pro cessing and no degradation of SNR at all.

6. REFERENCES

1) N. Benvenuto, G. Bertocci and W. R. Dauner : The 32-k b/s ADPCM Coding Standard, AT&T Tech. Jou., pp 12-21 65-5, Sep/Oct., 1986.
2) J. H. Chen, R. V. Cox, Yen-Chun Lin, N. Jayant, and M. J. Melchner: A Low-Delay Coder for the CCITT 16 kb/s Speech Coding Standard. IEEE Jou. Seletd. Areas in Comm. 10-5, pp 830-849, Jan. 1992.
3) A. Cichocki and R. Unbehauen: Simplified Neural Networks for Solv ing Linear Least Squares and Total Least Squares Problems in Real Time IEEE Trans. NN. pp 910-923, 5-6, Nov. 1994.