On Realizing the Predictor for the Waveform Coding of Speech Signals by using the Dual First Order Autocorrelation

—Using the Dual First Order Difference Values and the Sigma-Delta Technique—

쌍 1차 자기상관관계를 이용한 음성 파형부호화용 예측기의 구현

— 쌍 1차 차분값과 시그마-델타기법을 적용 —

(Misuk Lee*, Myungjin Bae*, Joohun Lee**)

이 미 숙*, 배 명 진*, 이 주 헌**

※본 논문은 체신부, 한국전기통신공사의 통신학술단체 육성지원금에 의해 이루어졌다.

ABSTRACT

The speech waveforms are highly correlated between the adjacent samples. One way of increasing the correlation in speech signals is to simply integrate the input signals prior to coding. The integrated values can be removed by conventional differentiation at the receiver. This emphasizes the low frequencies of speech signals and increases the correlation between adjacent samples. The above arrangement is called as a sigma-delta.

In this paper, we propose a new predictor which use such characteristics of sigma-delta. That is, we integrate input signals prior to coding and then, predict the present integrate sample by using two samples, one past and one next. The proposed predictor has higher mean prediction gain of 8.65dB than that of the CCITT-Recommendation ADPCM.

요 약

음성파형은 인근 표본값들 사이에 높은 상관관계를 나타낸다. 음성신호의 상관관계를 증가시키기 위한 한 방법으로는 부호화하기 전에 입력신호를 단순히 적분시키는 방법이 있다. 이 적분된 값들은 수신기에서 일반 미분기에 의해 제거될 수 있다. 이렇게 하면 음성신호의 저역주파수가 강조되고 인근 표본값의 자기 상관관계가 증가된다. 이런 과정을 시그마-델타 기법이라 한다.

이 논문에서는 그러한 시그마-델타의 특성을 사용하는 예측기를 새로이 제안한다. 즉, 부호화하기전에 입력신호를 적분하고 인근한 과거 및 미래의 두 표본을 사용하여 적분된 현재표본을 예측한다. 제안된 예측기는 CCITT-권고형 ADPCM의 평균 예측이득 보다 8.65db 높게 얻어졌다.

I. INTRODUCTION

The problem of coding speech signals for transmission or storage purposes has long been a subject of interest in speech research. Generally, speech coding algorithms can be classified into following three types:

In the waveform coding methods, the unnecessary redundancy in speech waveforms are reduced before it is transmitted through the

*Hoseo University
**Seoul National University

tansmission channel or stored in some storage medium. PCM, ADM and ADPCM belong to this type. Thanks to the improvement of the manufacturing techniques and algorithms of DSP(Digital Signal Processor), the ADPCM chip has realized with a bit rate of 32kbps. Also, the waveform coding methods can maintain the high quality and personality, because in the processing procedure, the vocal tract filter informations, which represent the meaning of message and the exicitation information that reflect the personality and feeling of a person, are not separated in two parts.

The source coding methods are very closely based on the speech production model. They separate the excitation information from the filter information in speech signals before these coding methods are realized. The methods that belong to this category are LPC, PARCOR, and LSP. These algorithms are very efficient in memory capacity because they have a transmission rate of 8~10 kbps.

The hybrid coding methods combine features from source coding and waveform coding and generally operate at a bit rate between two. MPLPC, RELP and VELP belong to this type [1-3]. This methods are not appropriate for synthesis-by-rule because it is difficult to change the source.

The area of speech waveform coding, especially, has received considerable attention in recent years due to increased efficiencies of coding implementions brought about by advances in large-scale integration of digital circuits. Therefore, if we want high quality in any application areas, then it is appropriate to use the waveform coding methods.

In this paper, we propose a new predictor which uses the characteristics of sigma-delta. That is, we integrate input signals prior to coding and then predict the integrated present sample by using two samples, one past and one next. In section Ⅱ, we shall review the basic principles of waveform coding and in section Ⅴ, we shall discuss the predictor which is proposed in this paper. The results for the proposed predictor, DPCM, CCITT-Recommendation ADPCM and the predictor which uses the dual first order difference values, respectively, are presented in section Ⅵ.

Ⅱ. ADPCM

In the areas of speech coding, we have two conflicting requirements : First, we want to achieve the lowest possible bit rate. Second, we want to achieve this with minimum loss in speech quality. But the information capacity required to transmit or store the digital representation capacity to transmit or store the digital representation is :

$$I = B \cdot F_s = \text{Bit rate in bits per second} \qquad (1)$$

Where, F_s is the sampling rate(i.e., samples/second) and B is the number of bit/sample. Thus, there are two ways in design a system that minmizes the transmission rate while maintaining a certain speech quality : First, choose the sampling rate to be many times the Nyquist rate and fix the bit rate per sample into one bit(0,1)(DM, ADM, etc.). Second, choose the sampling rate to be equal to the Nyquist rate and compress the bit rate per sample(DPCM, ADPCM, etc.).

In differential PCM(DPCM), the sampled speech signal is compared with a locally decoded version of the previous sample prior to quantization so that the transmitted signal is the quantized difference between samples. To minimize such differences, it is necessary to apply an adaptive technique to predictor or quantizer. Fig. 1 shows a block diagram of the ADPCM transmitter.

In the CCITT-Recommedation ADPCM, the adaptive predictive filter is a two-pole, six-zero filter used to determine the signal estimate. The sixth-order all-zero section helps to stabilize the filter and prevents it from drifting into oscil-

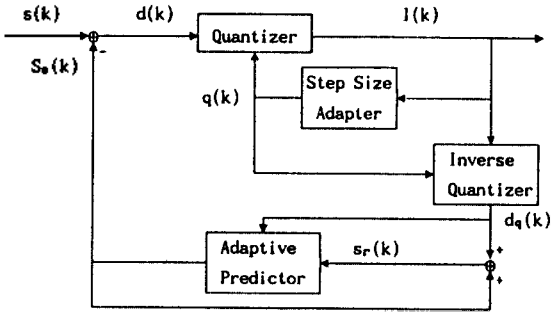


Fig. 1. ADPCM Transmitter Block Diagram.

lation.

In the predictors that were proposed so far, they use mainly the forward prediction methods, predict the present sample by using the linear combination of several past samples. But the correlation of speech signals decreases in proportion to time delay. Thus, it is not fitting to use the samples that are time-delayed by more than the fourth-order. It is the predictor which uses the dual first order difference values that predict the present sample by using the most adjacent two samples according to above correlation characteristics of speech signals.

Ⅲ. The Predictor Using the Dual First Order Difference Values

Genrally, speech signals are correlated among samples. It shows the variation from previous coarticulation state to present coarticulation state and changes according to time delay. Such correlation value is very high between adjacent samples but gradually decreases in proportion to time delay.

In the predictor which uses the dual first order difference values, we predict the present sample by using the most adjacent samples. This method does not require convergence time which appear generally in the linear predictor. And specifically, in the voiced segment, the predictor which uses the dual first order difference values has higher prediction gain than that of the CCITT-Recommendation ADPCM. However, this predictor is somewhat vulnerable to noise.

Ⅳ. Consideration for the Prediction Error

In the predictor using the dual first order difference values, the predicted sample $\hat{s}(n)$ for the present sample obtained by averaging the past sample $s(n-1)$ and next sample $s(n+1)$ in the original singal.

Where, prediction error $e_p(n)$ is shown as follows :

$$\begin{aligned} e_p(n) &= s(n) - \hat{s}(n) \\ &= s(n) - \{s(n+1)+s(n-1)\}/2 \end{aligned} \tag{2}$$

However, at the receiver, to reconstruct the present sample by using the above prediction method, we have to know the value of the next sample. Thus we use the present sample which is compensated with the difference value between reconstructed next sample and original next sample as a next sample. In the coding procedure, prediction error $e_p(n)$ for the present sample compensated with difference $e_n(n)$ of next sample is,

$$\begin{aligned} d(n) &= e_p(n)+e_n(n) \\ &= s(n) - \{s(n+1)+s(n-1)\}/2 \\ &\quad + \{s^*(n+1) - s(n+1)\}/2 \end{aligned} \tag{3}$$

The compensated prediction error $d(n)$ coded by the coder, thus add the error $e_q(n)$ for the quantization to $d(n)$. In the synthesis procedure, prediction error which is coded in the transmitter is decoded at the receiver and also adds the error $e_i(n)$ due to the decoder to $d(n)$ as follows :

$$\hat{d}(n) = d(n)+e_q(n)+e_i(n) \tag{4}$$

At the receiver, the next sample $s^*(n+1)$ will be reconstructed by using the two past samples, $s^*(n)$ and $s^*(n-1)$, and the prediction error for the present sample. It is shown as follows :

$$\begin{aligned} s^*(n+1) &= 2s^*(n) - s^*(n-1) - 2\hat{d}(n) \\ &= 2s^*(n) - s^*(n-1) - 2\{e_q(n)+e_i(n)\} - 2s(n) \end{aligned}$$

$$+s(n+1)+s(n-1)+s(n+1)-s^*(n+1) \quad (5)$$

In Eq.(5), if the errors which occur in quantization or decoding can be negligible then the reconstructed sample will be nearly the same with original sample. Thus the reconstructed next sample can be represented as follows :

$$s^*(n+1)=s(n+1)-2\{e_q(n)+e_i(n)\} \quad (6)$$

However, the reconstructed signals are sensitive to the error of the coder and decoder twices as much. In this paper, we attempt to decrease the sensitivity of this predictor to noise by using the sigma-delta technique. That is, we integrate input signals prior to coding so as to increase correlation values between adjacent samples.

V. The Predictor Using the Dual First Order Difference Values and the Sigma-Delta Technique

The correlation between sample and sample gradually decreases as the time delay increases. Also, this characteristics of speech signals is shown in the voiced segments which have relatively high correlation values. Thus, it is more appropriate to use adjacent two samples (i.e, one past and the other next) than to use simply time delayed samples. In the predictor using correlation characteristics of speech signals, if we can increase the correlation between samples, then we can obtain higher prediction gain.

Table.1 The Correlation Value of Speech Signals.

order	female		male	
	Dual-ACF	Σ-ACF	Dual-ACF	Σ-ACF
1	95.2	96.5	96.1	96.7

Table.1 shows the mean correlation values between the averaged value of the one past sample and the other next and the original present sample. The speech data used in this experiment comprise 5 sentences(spoken by 3 males and 2 females) which include nasal, voiced, unvoiced and silence segments. Dual-ACF represents the correlation values for the original signals and Σ-ACF represents the correlation values for the integrated speech signals.

As is found in table.1, the integrated speech signals have higher correlation values than the original signals(i.e., not integrated singals). Thus in the proposed predictor, we integrate input signals and then predict the present sample by using the one past sample and one next sample. At the receiver, we implement the differentiation to reconstruct the original signals. Eq.7 represents the integration for input signals at the transmitter and Eq.8 represents the differentiation for the integrated signal at the receiver. This is illustrated in Fig. 2. Where, 0.5 factor in the Σ-block is used to prevent the divergence which can occure according to dc offset values. Thus in the d-block, the added 2 factor corresponds to 0.5 in Σ-block.

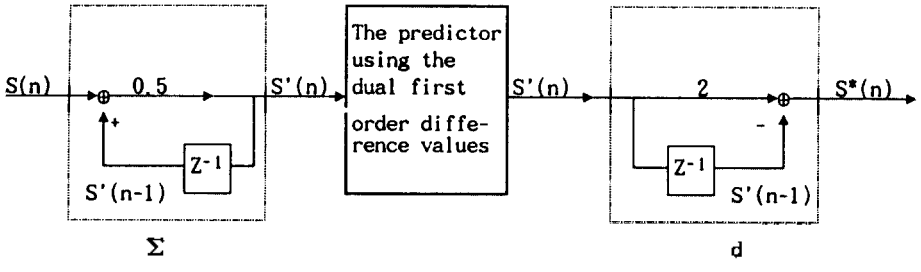


Fig.2. Block Diagram for total Processing Procedure.

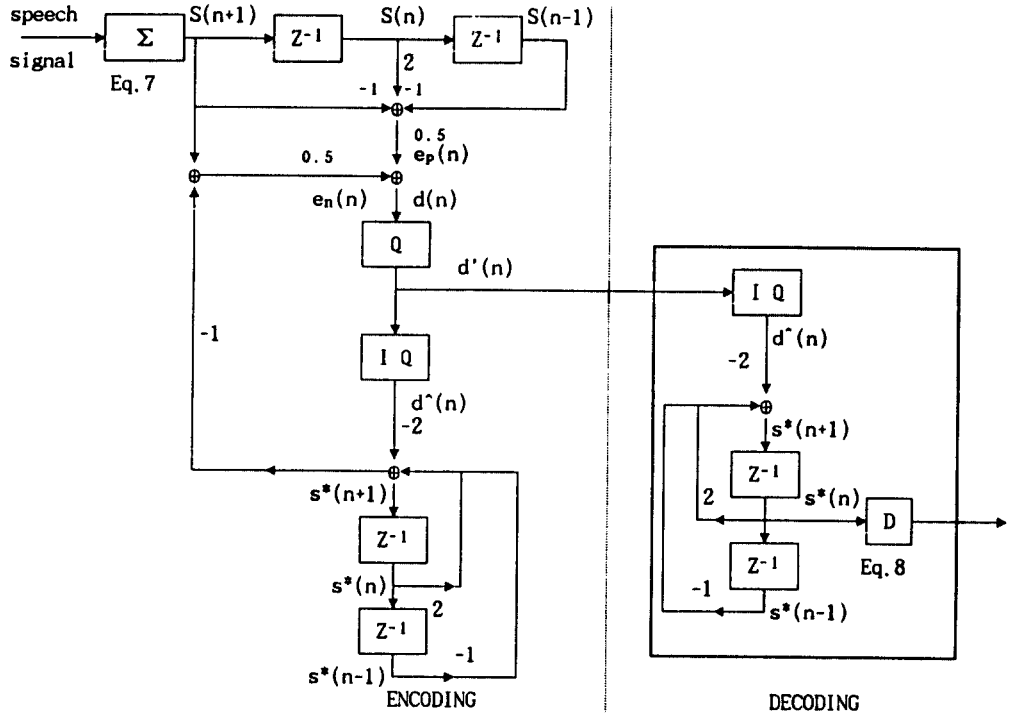


Fig.3. CODEC by using the Dual First Order Difference and the Sigma-Delta Technique

$$s'(n)=\{s'(n-1)+s(n)\}/2 \quad (7)$$

$$s^*(n)=2\times s'(n)-s'(n-1) \quad (8)$$

$$=2\times\frac{s'(n-1)+s(n)}{2}-s'(n-1)$$

$$=s(n)$$

Fig.3 shows the block diagram for the predictor which is proposed in this paper.

VI. Experimental Results

The proposed algorithm has been implemented on an IBM PC/AT with the 12-bit A/D converter. The sampling frequency is 8KHz. And no special window for preweighting signals was used. The speech data are comprised of 5 speakers, 3 males and 2 females. The following sentences were spoken.

1) /HOSEODAE JUNJAKONGHAKWA UMSEONG SINHOCHURI YUNGU/
2) /KAMSAHAMNIDA/
3) /JIGUMGEOSIN JUNHWANUN/
4) /JESUNIMKESEO CHUNJICHANGJOWI KIOHUNWL MALSUMHASEOSSDA/
5) /INSUNE KOMAGE CHUNJAA SONYUNWL JOAHANDA/

Each sentence was recored with little background noise. We chose the frame length with 256 points, and processed each frame as shown in Fig.3. To obtain $s^*(n)$, we averaged one past sample and one next sample and added the error to predicted present sample. Then using the relatio of $s(n)$ and $s^*(n-1)$ obtained the $s^\wedge(n+1)$.

We experimented for the clean speech and also Gaussian-noise corrupted speech at 20, 12, 6, 0dB SNR.

Fig.4 illustrates the time-varying prediction gain achieved for the utterance(1). It is clear that the proposed predictor in this paper attains higher gain than others, particularly in the voiced segments. Prediction gains for the other inputs

show similar behavior.

Tables.2 and 3 illustrate the performance scores for each of the four predictors at four signal-to-noise ratios. In these tables, we can observe the fact that the proposed predictor has the best performance score at all four SNR's.

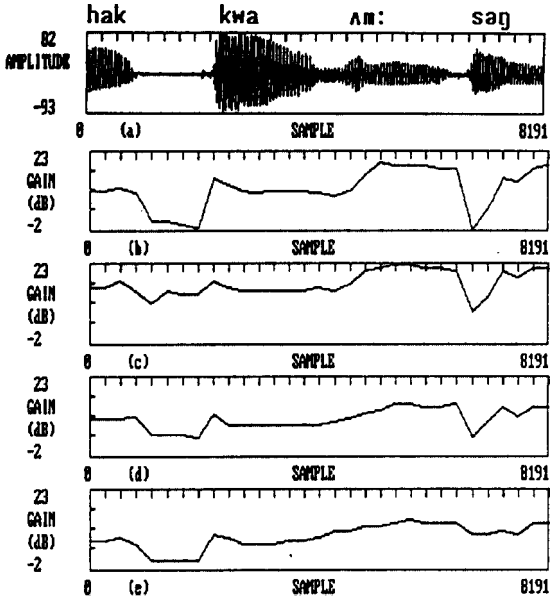


Fig.4. Time-varying prediction gains for four prediction algorithm : (a) Speech Waveform : (b) The Predictor using the Dual First Order Difference Values : (c) The Proposed Predictor : (d) The Predictor of DPCM : (e) The Predictor of CCITT-Recommendation ADPCM.

Table.2. Performance Scores for Four Predictors

	DUAL-DFP (dB)	Σ-DFP (dB)	DPCM (dB)	CCITT (dB)	Improvement for the CCITT
ORIGINAL	15.27	19.42	12.12	9.20	10.22
20dB	8.65	15.74	10.49	6.06	9.68
12dB	3.03	11.11	6.90	4.33	6.78
6dB	-1.58	6.89	2.89	2.11	4.78
0dB	-5.08	3.41	-0.43	0.43	2.98

Table.3. Performance Scores for Four Predictors

	DUAL-DFP (dB)	Σ-DFP (dB)	DPCM (dB)	CCITT (dB)	Imp men the C
ORIGINAL	11.47	17.24	9.98	9.6	8.
20dB	7.87	15.37	9.02	6.70	8.
12dB	3.29	11.87	6.57	4.38	7.
6dB	-1.28	7.00	2.95	1.94	5.
0dB	-4.89	3.45	-0.30	0.43	3.

Ⅶ. Conclusions

The problem of coding speech signals for transmission or storage purposes has long been a subject of interest in speech research. In the predictor proposed untill now, it is difficult to mix several coding methods because of the convergence speed of the predictor. And the CCITT-Recommendation ADPCM(G.721) has a very complex scheme. The predictor which was proposed by our laboratory in 1991 is sensitive to noise.

Thus in this paper, we used the sigma-delta technique to reduce the sensitivity to noise. That is, we have integrated the input signal and predicted the present sample by using the one past sample and one next sample. At the receiver, we applied differentiation to reconstruct the original signal.

The proposed predictor does not require convergence time and is robust in noise. Also the proposed predictor has a higher mean prediction gain of 8.65dB than that of the CCITT-Recommendation ADPCM.

REFFERENCES

1. L.R. Rabiner & R.w. Schafer, Digital processing of speech signals, Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 1978.
2. Doglas O'Shaughnessy, Speech Communication : Human and Machnine, Addition-Wsley Publishing Company 1987.
3. P.E.Papamichalis, Practical Approaches to Speech Coding, Prentice-Hall, Inc., Englweod Cliffs, New Jersey, 1987.

4. Jan. P. van Hemert, "Automatic Segmentation of Speech", IEEE Trans. Acoust., Speech and Signal Proc. vol. ASSP-39, No.4. pp. 1008−1012. Apr.1991.
5. Jayant & Noll, Digital Coding of Waveforms, Prentice-Hall, Alan V.Oppenheim, series editor.

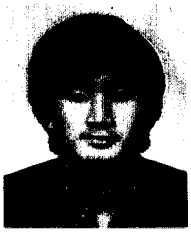
▲Mi Suk Lee

was born in Kongju, Korea, on March 15, 1968. She received the B.S. degree in electronics engineering from HoSeo University, Korea in 1991. She is currently enrolled in a M. S. degree at the Hoseo University. Her current research interests are in speech signal processing and its application.

▲Joohun Lee

was born in Seoul, Korea, on June 19, 1964.
He received the B.S., M.S., degree in electronics engineering from seoul National University, Seoul, in 1988 and 1990, respectively. He is currently working toward the Ph.D degree at seoul National University. His current areas of research are in the communication theory, digital signal processing and the statistical communication theory.

▲Myungjin Bae



was born in Kyungbook-do, on May 20, 1956. He received the B.S. degrees in electronics engineering from Soongsil University, Seoul, in 1981. He also received the M.S. and Ph.D. degree in electronics engineering from Seoul National University, Seoul, in 1983 and 1991, respectively.

Since 1986, he has been with the department of Electronics Engineering, Hoseo University, Chunan-si, where he is currently an Assistant Professor. His research interests include speech signal processing, adaptive signal processing, and communication system.