KAK 필터를 이용한 잡음이 섞인 음성의 음질향상

On the Use of a KAK Filter for Enhancement of Noisy Speech

* 조 동 호 (Cho, D. H.)
** 유 득 수 (You, D. S.)
*** 은 종 관 (Un, C. K.)

요 약

본 논문에서는 광대역 또는 협대역잡음이 섞인 음성의 음질을 개선하기 위해서 Kak필터를 사용하는 방법을 제안한다. Kak필터는 그 구조가 간단하지만, 잡음이 섞인 음성의 음질을 개선하는데 있어서 객관적인 음질척도로 볼 때 spectral subtraction 방법과 성능이 비슷하다. 또한 귀로 들어봐도 Kak필터를 사용한 경우와 spectral subtraction 방법을 이용한 경우의 개선된 음질이 거의 비슷하다. 그런데 이 Kak필터는 그 구조가 다른 기존방법보다 훨씬 간단하며, 다른 음질개선 알고리즘과는 달리 음성과 묵음의 판별이 필요하지 않다. 또한 Kak필터는 ADPCM과 같은 파형 부호화기와 결합하는 것이 용이하다. 따라서 깨끗한 음성뿐만 아니라 잡음이 섞인 음성을 부호화하는데 있어서 제안한 Kak필터를 ADPCM과 같은 파형 부호화기에 결합하여 사용하는 것이 적합하다.

ABSTRACT

In this paper an algorithm for enhancing noisy speech corrupted by white or colored noise using the Kak filter is proposed. Although the Kak filter is simple in its structure, it yields a performance improvement that is comparable to the spectral subtraction method in enhancing noisy speech. According to our informal listening test, the enhanced speech by the Kak filter has the quality comparable to that obtained by the spectral subtraction method. The enhancing algorithm is much less complex than existing

*, *** 한국과학기술원 통신공학실, ** 한국전자통신연구소

algorithms. Moreover, unlike other algorithms, it does not require speech/silence discrimination. Also, it is easy to combine with waveform coders such as adaptive differential pulse code modulation (ADPCM). Hence, the proposed enhancing algorithm incorporated in ADPCM is ideally suitable for coding noisy speech as well as clean speech.

I. INTRODUCTION

In general, the performance of a speech compression system degrades rapidly when there exist additive noise and other distortions. For this reason, considerable effort is currently being made for the development of a speech coder that is robust to additive noise and acoustical distortion. There are two basic approaches in designing a robust speech coder, especially at a low rate. One approach is to use a preprocessor to enhance the degraded speech before processing by the bandwidth compression system [1]. Another approach is to incorporate an enhancement scheme into a speech coder [2].

The perceptual aspects of speech are considerably complicated, and are not well understood. From the view point of speech perception, it is known that short-time spectral magnitude is more important than phase. The enhancement systems based on this concept are the spectral subtraction technique and the optimum Wiener filtering technique [3],[4].

Unlike the enhancement systems utilizing perceptual attributes, some enhancement schemes are based on the utilization of the harmonic structure of voiced speech. In this scheme the short-time periodicity of speech during voiced interval is used to enhance noisy speech. One example is the comb filtering approach in which components of speech spectrum only at harmonic frequencies are extracted [5].

Another approach to speech enhancement is to estimate parameters of a speech prodction model. Enhanced speech is obtained by passing the noisy speech through an analysis-synthesis system formed by those parameters. So far, several approaches based on all-pole or pole-zero modeling have been proposed for enhancement of noisy speech using a speech production model [6],[7].

In addition, to track and reject narrow-band noise signal, the use of a time-domain filter has been investigated. It may be formed from the inverse transform of the inverse of the estimated noise spectrum. This filter can be implemented using a time-domain noise suppression filter that is adapted segmentally based on samples of background noise [8].

In this paper, a new enhancement technique using a Kak filter is proposed for enhancing noisy speech corrupted by white or colored noise. The Kak filter, which has been used primarily for image processing [9], is simple and effective for speech enhancement, and it can easily be incorporated in existing adaptive differential pulse code modulation (ADPCM) systems for speech coding. The performance of the proposed method is considered analytically, and its performance is compared to that of the spectral subtraction method which is known to be an effective method for speech enhancement.

Following this introduction, the Kak filter algorithm for enhancement of noisy speech is described in Section II. Also, the performance improvement by the Kak filter is analyzed. In Section III, the enhancement of noisy speech by the Kak filter is explained. In Section IV, computer simulation for enhancement of noisy speech corrupted by white or colored noise by the Kak filter algorithm is done. In addition, the use of thw Kak filter algorithm in an ADPCM system is studied, and the performance of the combined system is investigated. Finally, conclusions are made in Section V.

II. THE KAK FILTER ALGORITHM AND ITS PERFORMANCE

A. Fixed Kak Filter Algorithm

Let us first consider a fixed Kak filter algorithm for enhancing noisy speech. Let $\{x(n)\}$ be samples of input speech that may be represented by an m-th order autoregressive process generated by $\{z(n)\}$ which is a Gaussian process. It is expressed as

$$x(n) = \sum_{l=1}^{m} a_l x(n-k) + z(n), \quad \text{for } n=1, 2, \cdots,$$
$$= 0, \qquad \text{otherwise} \quad (1)$$

where $\{a_l\}$ are prediction coefficients, and z(n) has zero mean and variance of σ^2. When the speech is corrupted by additive white Gaussian noise w(n), noisy speech, $\{y(n)\}$, may be expressed as

$$y(n) = x(n) + w(n) \quad (2)$$

where w(n) is assumed to be uncorrelated with input speech. The Kak filter is a point-to-point recursive estimator based on the DPCM model given by (1) and observations such as noisy speech of (2) [10]. However, this filter is not a one-step predictor. The n-th estimator output $\hat{x}(n)$ is expressed

$$\hat{x}(n) = (1-\eta_n)\,\tilde{x}(n) + \eta_n y(n) \quad (3)$$

with

$$\tilde{x}(n) = \sum_{l=1}^{m} a_l \hat{x}(n-l) \quad (4)$$

where $\tilde{x}(n)$ is the n-th predicted value based on the previous reconstruction values, and η_n is a filter gain of the n-th sample. A block diagram of the fixed Kak filter is shown in Fig. 1. We can note from this figure that the Kak filter becomes a conventional DPCM if the multiplier is replaced by a quantizer. This indicates that the Kak filter can easily be combined with a DPCM coder (This aspect will be considered in Section III.)

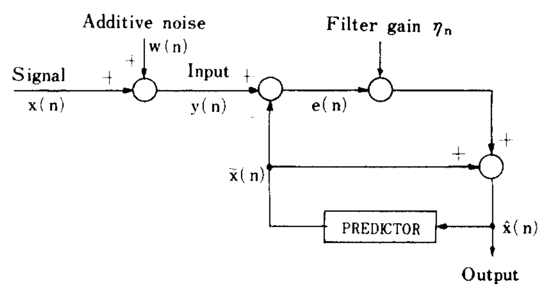

Fig. 1 Block diagram of a fixed Kak filter.

To realize the Kak filter, $\{a_l\}$ and η_n must be determined. $\{a_l\}$ are obtained by solving a set of normal equations composed of autocorrelation functions. Also, if we wish $\hat{x}(n)$ to be a minimum mean-squared error (MMSE)estimate, η_n can be calculated by applying the orthogonality principle as the following:

$$E[\{x(n) - \hat{x}(n)\}y(n)] = 0, \quad (5)$$

where $E[\cdot]$ is the expected value. Using (2), (3), (4) and (5), it may be shown that η_n is given by

$$\eta_n = \frac{\sigma_x^2 - E[\tilde{x}(n)x(n)]}{\sigma_x^2 - E[\tilde{x}(n)\ x(n)] + \sigma_w^2}. \tag{6}$$

where σ_x^2 and σ_w^2 are variances of x(n) and w(n), respectively. From (6), one can see that, when signal-to-noise ratio (SNR) is large, i.e., $\sigma_x^2 \gg \sigma_w^2$, the gain η_n approaches to unity. This implies that the estimate $\hat{x}(n)$ is based almost on the current observation y(n) when the SNR is large. On the other hand, in the case of low SNR, the estimate depends almost entirely on the predicted values based on the previously reconstructed values.

Next, let us consider the performance improvement by the Kak filter [11]. In general, the performance improvement by enhancement, SNR_{gain}, is expressed as

$$SNR_{gain} \triangle \frac{\text{SNR of enhanced speech}}{\text{SNR of noisy speech}} = \frac{\dfrac{E[x^2]}{E[(x-\hat{x})^2]}}{\dfrac{E[x^2]}{E[w^2]}} \tag{7}$$

Using (2), (3) and (7), we can write SNR_{gain} as

$$SNR_{gain} = \frac{\sigma_w^2}{(1-\eta_n)^2 E[\{x(n) - \tilde{x}(n)\}^2] + \eta_n^2 \sigma_w^2} \tag{8}$$

One can see from (8) that there is no improvement when η_n is equal to 1. However, since η_n is in general less than 1, one can have a performance improvement if the variance of the difference between clean speech and predicted value is less than that of the noise.

B. Adaptive Kak Filter Algorithm

It is known that the autoregressive moving average (ARMA) filter is an optimum innovation filter for the speech model of (1). Here, we consider briefly the ARMA filter, and show that an adaptive Kak filter may be regarded as a suboptimal version of the ARMA filter [12], [13]. Let e(n) be the difference between the input y(n) and predicted value $\tilde{x}(n)$ given by

$$\tilde{x}(n) = \sum_{l=1}^{m} a_l y(n-l) + \sum_{l=1}^{m} b_l e(n-l), \tag{9}$$

where AR filter coefficients $\{a_l\}$ are obtained by solving a set of normal equations composed of autocorrelation functions, and MA filter coefficients $\{b_l\}$ are calculated using the truncation property [13]. Then, the output $\hat{x}(n)$ of the ARMA filter is given by

$$\hat{x}(n) = y(n) - \frac{\sigma_w^2}{\sigma_e^2} e(n), \tag{10}$$

where σ_e^2 is the variance of e(n). To compare the algorithm of the Kak filter with that of the ARMA filter, we can rewrite (9) and (10) as

$$\tilde{x}(n) = \sum_{l=1}^{m} a_l \hat{x}(n-l) + \sum_{l=1}^{m} d_l e(n-l) \tag{11}$$

$$\hat{x}(n) = (1-\eta_a)\tilde{x}(n) + \eta_a y(n) \tag{12}$$

where

$$e(n) \triangle y(n) - \tilde{x}(n), \tag{13}$$

$$\eta_a \triangle 1 - \frac{\sigma_w^2}{\sigma_e^2}, \tag{14}$$

$$d_l \triangle b_l + \frac{\sigma_w^2}{\sigma_e^2} a_l. \tag{15}$$

One can note from (3), (4), (11) and (12) that the Kak filter is a suboptimal version of the ARMA filter since the MA part of the ARMA is calculated approximately using the AR part. Also, note that the filter gain may be updated according to (14) sample by sample or block by block. Unlike the fixed Kak filter algorithm, the filter gain η_a is updated adaptively by using the variances of noise component and quantization noise. A block diagram of the adaptive Kak filter is shown in Fig. 2.

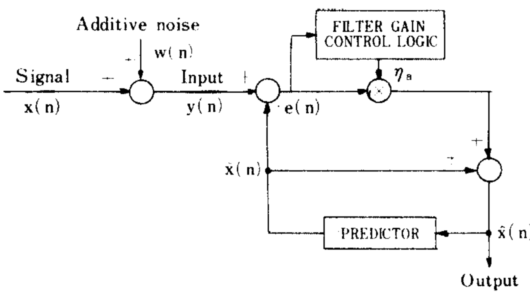

Fig. 2 Block diagram of an adaptive Kak filter.

III. STRUCTURES OF ENHANCEMENT SYSTEMS USING A KAK FILTER

Here we investigate two different structures, one having the cascaded form in which the Kak filter and a speech coder are connected in series, and the other having the combined form in which the Kak filter algorithm is incorporated into a speech coder such as DPCM or ADPCM.

A. Cascaded Structure

The cascaded structure is shown in Fig. 3. Like various conventional enhancement algorithms, preprocessing for enhancement of noisy speech is done before speech is coded. In this case, either fixed or adaptive Kak filter algorithm discussed in Section II can be used. As one can see, the structure of the Kak filter is simple. Furthermore, unlike the conventional enhancement algorithms, the fixed Kak filter does not require speech/silence discrimination which takes a considerable amount of computations.

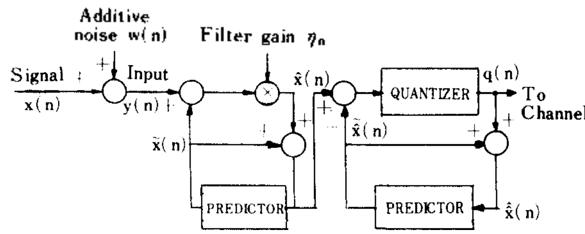

Fig. 3 A cascaded structure of the Kak filter and a DPCM coder.

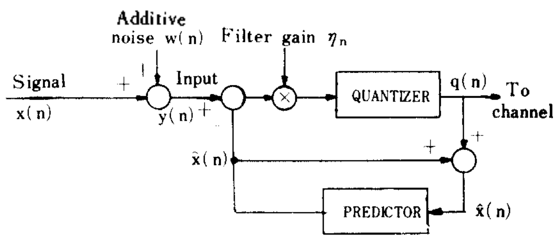

Fig. 4 A combined structure of the Kak filter and a DPCM coder.

B. Combined Structure

In the case that the Kak filter is applied to a waveform coder such as DPCM and ADPCM, a combined structure is possible. As discussed in Section II, the structure of the Kak filter is similar to that of a DPCM coder. That is, if the quantizer of the DPCM coder has an infinite number of levels, the coder becomes the same as the Kak filter. The combined structure is shown in Fig. 4. In this structure, enhancement and coding are done simultaneously by adjusting the filter gain prior to quantization. The value of the filter gain is chosen such that the SNR of enhanced speech may be maximized and updated adaptively using the variances of noise component and quantization noise. For this combined structure, either fixed or adaptive Kak filter can again be used. Also, since enhancement and coding are done simultaneously, delay does not occur. Moreover, in comparison to the cascaded form, the combined structure is very simple and easy to implement. Besides, the fixed Kak filter of the combined form does not require speech/silence discrimination. Hence, in view of the enhancement and computational complexity it would be desirable to use the fixed Kak filter of the combined form for enhancement in ADPCM coding of noisy speech.

IV. COMPUTER SIMULATION RESULTS AND DISCUSSION

A. Performance Improvement

We now investigate the performance improvement by the Kak filter algorithm in enhancing noisy speech, and compare it to that of the spectral subtraction method by SNR and segmented SNR.[1] As an input to these systems, real speech bandlimited to 3.4 kHz and sampled at 8 kHz was used. To obtain noisy speech corrupted by white noise, we generated white Gaussian noise using a random number generation program. This noise was then processed by a low-pass filter whose 3 dB cutoff frequency was 3.4 kHz, and added to clean speech. Also, to obtain noisy speech corrupted by colored noise, white Gaussian noise was processed by a band-pass filter whose frequency response is shown in Fig. 5. It was then added to clean speech. Note that narrow-band ridges of the spectrum in Fig. 5 correspond to the fundamental (1550 Hz) and first harmonic (3100 Hz) narrow-band noise of the helicopter engine [3].

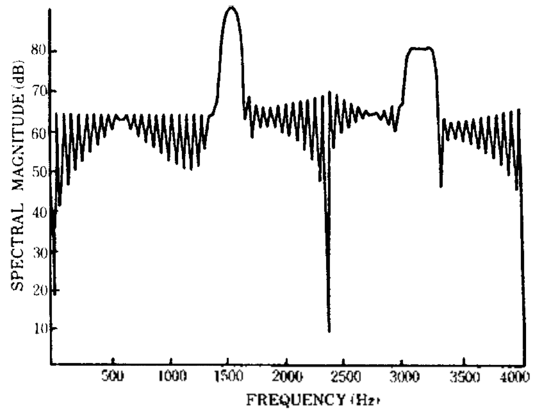

Fig. 5 The frequency response of bandpass filter to generate colored noise.

In our simulation, we first evaluated the performance of the Kak filter in cascaded with a DPCM coder (see Fig. 3). A comparison of the performance improvement resulting from the use of the Kak filter and also the spectral subtraction method is shown in Table I. It is seen that the performance of the Kak filtering method is similar to that of the spectral subtraction method in improving noisy speech. The performance improvement is in the range of 1 to 3 dB in SNR_{SEG} when the SNR of input speech

---

[1] Segmented SNR is defined as

$$SNR_{SEG}(dB) = \frac{10}{N} \sum_{i=1}^{N} \log_{10} (1 + SNR_i)$$

where SNR_i is the SNR of i-th frame composed of M speech samples and N is the total number of frames.

Table I. Performance improvement in SNR and segmented SNR resulting from enhancement by the Kak filtering and the spectrum subtraction method for noisy speech corrupted by white noise.

Enhancement (dB) / Noisy speech (dB)	Kak filter		Spectral subtraction
	Fixed	Adaptive	
0 dB (0.61)	3.65 (2.61)	4.25 (2.78)	4.39 (3.12)
5 dB (3.73)	7.56 (5.14)	8.60 (5.52)	9.02 (6.08)
10 dB (6.82)	11.21 (7.43)	12.48 (8.04)	12.58 (8.46–

Note: The values in parenthesis are segmented SNR's

is in the range of 0 to 10 dB. Also, it is seen that the adaptive ak filter yields slightly better performance than the fixed one. As one can expect, the performance improvement decreases as the SNR of input speech becomes higher. The reasons are thought to be due to the non-stationary characteristics of speech and also because there is little room for improving noisy speech with high SNR. Fig. 6 shows waveforms of clean speech, 5 dB noisy speech and enhanced speech by the Kak filter. We can see from this figure that the adaptive Kak filter is superior to the fixed Kak filter particularly in the unvoiced portion, although the performance gain in SNR_{SEG} of the adaptive Kak filter over the fixed one is relatively small.

The performances of the Kak filter and the spectral subtraction method in enhancing noisy speech corrupted by colored noise are shown

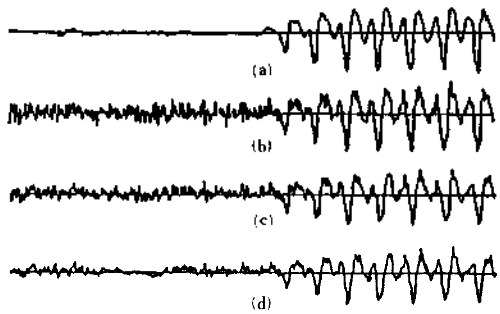

Fig. 6 Waveforms of original, 5 dB noisy and enhanced speeches
(a) Original speech
(b) 5 dB noisy speech corrupted by white noise
(c) Enhanced speech by fixed Kak filter
(d) Enhanced speech by adaptive Kak filter.

in Table II. It is seen that the performance of the adaptive Kak filter is again slightly better than that of the fixed Kak filter. Also, note that the performance of the fixed or adaptive

Table II. Performance improvement in SNR and segmented SNR resulting from enhancement by the Kak filtering and spectrum subtraction method for noisy speech corrupted by colored noise.

Enhancement (dB) / Noisy speech (dB)	Kak filter		Spectral subtraction
	Fixed	Adaptive	
0 dB (0.81)	5.04 (3.4)	4.48 (3.17)	3.64 (3.25)
5 dB (4.03)	8.5 (5.82)	8.67 (5.86)	8.4 (6.44)
10 dB (7.44)	11.90 (8.18)	12.31 (8.62)	12.55 (9.29)

Note: The values in parenthesis are segmented SNR's

Kak filter is similar to that of the spectral subtraction method. Moreover, according to our informal listening tests, it appears that the performance of the Kak filtering method is as good as that of the spectral subtraction method.

As discussed in Section III, the fixed or adaptive Kak filter may be incorporated in a DPCM system for enhancement in waveform coding of noisy as well as clean speech. In our simulation we used a DPCM coder with a gamma quantizer and a one tap-predictor, and also the CCITT[2] standard ADPCM coder. The performance improvements in the reconstructed speech by the fixed or adaptive Kak filter in the cascaded or combined form are shown in Table III for noisy speech corrupted by white noise. For noisy speech with colored noise, a similar performance gain could be obtained. We can see from Table III that, for noisy speech of 0, 5

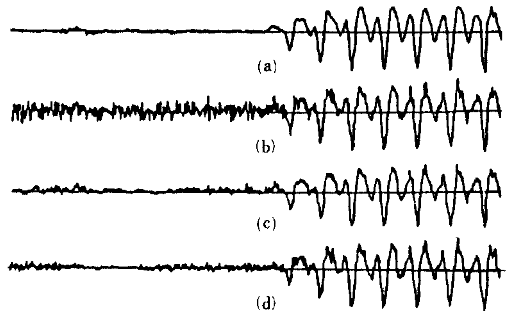

Fig. 7 Waveforms of clean, 5 dB noisy and enhanced speeches coded by CCITT-standrrd ADPCM at 32 kbits/s
(a) Original speech
(b) CCITT ADPCM-coded 5 dB noisy speech corrupted by white noise
(c) Enhanced waveform by the cascaded form of CCITT-ADPCM and adaptive Kak filter
(d) Enhanced waveform by the combined form of CCITT-ADPCM and adaptive Kak filter.

Table III. Performance improvement in segmented SNR resulting from enhancement by the cascaded or combined form of the Kak filter and a waveform coder for noisy speech corrupted by white noise.

Noisy speech	Enhancement / Coding / Form	Fixed Kak filter		Adaptive Kak filter	
		DPCM	CCITT ADPCM	DPCM	CCITT ADPCM
0.61 dB	Cascaded	2.53	2.5	2.77	2.66
	Combined	2.51	2.4	2.71	2.56
3.73 dB	Cascaded	5.15	4.96	5.53	5.31
	Combined	5.09	4.75	5.46	5.27
6.82 dB	Cascaded	7.49	7.19	8.05	7.71
	Combined	7.48	7.1	7.9	7.7

[2] International Telegraph and Telephone Consultative Committee

and 10 dB, the performances of the DPCM and the CCITT-standard ADPCM are almost the same. The gain resulting from the use of an adaptive scheme appears to exist only for clean speech. In addition, we can note that the cascaded form has a little better performance than the combined form. Fig. 7 shows noisy speech, clean speech and enhanced speech by the adaptive Kak filter applid to the CCITT-standard ADPCM in the cascaded or combined form for noisy speech corrupted by white noise. It is noted that there exists little difference between the cascaded and the combined forms.

V. CONCLUSIONS

In this paper, we have studied a Kak filter algorithm for enhancing noisy speech corrupted by white or colored noise. It has been shown that the Kak filter yields almost the same performance as the spectral subtraction method in enhancing noisy speech, yet it is much simpler in complexity. Furthermore, unlike the conventional enhancement algorithms, it does not require speech/silence detection. The Kak filter can easily be combined with a DPCM coder (e.g., the CCITT-standard ADPCM system) with little increase in the system complexity. The Kak filtering method yields the performance comparable to the spectral subtraction method in enhancing speech corrupted by white or colored noise.

REFERENCES

1. J.S. Lim and A.V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," Proc. IEEE, vol. 67, pp. 1586-1604, Dec. 1979.
2. J.C. Lee, C.K. Un and D.H. Cho, "A frequency-weighted block LMS algorithm and its application to speech processing," Proc. of IEEE, vol. 73, pp. 1137-1138, June 1985.
3. S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-27, pp. 113-120, Apr. 1979.
4. W.K. Pratt, "Generalized Wiener filtering computation techniques," IEEE Trans. Compt., vol. C-21, pp. 636-641, July 1972.
5. T.W. Parsons, "Separation of speech from interfering speech by means of harmonic selection," J. Acoust. Soc. Amer., vol. 60, pp. 911-919, Oct. 1976.
6. J.S. Lim and A.V. Oppenheim, "All-pole modeling of degraded speech," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-26, pp. 197-210, June 1978.
7. K.H. Song and C.K. Un, "Pole-Zero modeling of speech based on high-order pole model fitting and decomposition method," IEEE Trans. on Acoust., Speech and Signal Processing, vol. ASSP-31, No. 6, pp. 1556-1565, Dec. 1983.
8 L. Hoy et al., "Noise suppression methods for speech applications," in Proc. 1983 IEEE Int. Conf. Acoust., Speech, Signal Processing, pp. 1133-1136, Apr. 1983.
9. A. Rosenfeld and A.C. Kak, *Digital Picture Processing*, Academic press, New York, pp. 203-255, 1976.
10. J. Biemond and J.J. Gerbrands, "Comparison of some two-dimensional recursive point-to-point estimators based on a DPCM image model," IEEE Trans. Syst. Man Cyber., vol. SMC-10, pp. 929-936, Dec. 1980.
11. T. Katayama, "Restoration of noisy images using a two-dimensional linear model," IEEE Trans. Syst. Man Cyber., vol. SMC-9, pp. 711-717, Nov. 1979.
12. A.S. Willsky, "Relationships between digital signal processing and control and estimation theory," Proc. IEEE, vol. 66, pp. 996-1017, Sept. 1978.
13. M.G. Strintzis, "Dynamic representation and recursive estimation of cycle and two-dimensional processes," IEEE Trans. Auto. Control, vol. AC-23, pp. 801-808, Oct. 1978.
14. Y.H. Yum and S.B. Park, "Optimum recursive

filtering of noisy two-dimensional data with sequential parameter identification," IEEE Trans. Patt. Anal. Machine Intell., vol. PAMI-5, pp. 337-344, May 1983.

15. P. Noll, "A comparative study of various quantization schemes for speech encoding," Bell Syst. Tech. J., vol. 54, pp. 1597-1614, Nov. 1975.

16. J.D. Gibson, "Adaptive prediction in speech differential encoding systems," Proc. IEEE, vol. 68, pp. 488-525, Apr. 1980.

17. CCITT Report, Study Group XVIII, Temporary Document 18, Draft Recommendation G. 721, Geneva, Nov. 21-25, 1983.