論文 94-19-8-17

트리 코딩에서 전송에러에 강한 역방향 적응 피치 예측

正會員 李 寅 誠*

Robust Backward Adaptive Pitch Prediction for Tree Coding

Insung Lee* *Regular Member*

要 約

피치 예측기는 강인한 트리 부호화기에서 가장 중요한 부분중에 하나이다. 피치 예측기는 역방향으로 블럭 적용 방법과 회귀적인 방법이 결합되어 구성되어진다. 부호화기의 전송에러에 대한 성능을 개선하고 입력 음성의 피치주기의 변화를 추적하기 위해 피치 예측기의 스무더를 부가하는 방법을 제시한다. 3개의 탭을 갖는 스무더는 고정된 계수를 가지거나 피치 합성기의 출력신호의 자기상관 함수에 따라 변화되는 가변계수를 가질수 있다. 피치 예측기에 스무더의 부가는 한 블럭내에서의 피치주기의 변화를 추적할 수 있고 채널에러에 대한 영향도 줄일수 있다.

ABSTRACT

The pitch predictor is one of the most important part for the robust tree coder. The hybrid backward pitch adapation which is a combination of a block adaptation and a recursive adaptation is used for the pitch predictor. In order to improve the error performance and track the pitch period change of the input speech, it is proposed to smooth the input of the pitch predictor. The smoother with three taps can have fixed coefficients or variable coefficients depending on the estimated autocorrelation function of the output of the pitch synthesizer. The inclusion of a variable smoother can track the pitch period change within a block and reduce the effect of channel errors.

Ⅰ. INTRODUCTION

Speech signals can be effciently modeled with two slowly time-varying linear prediction filters that produce the spectral envelope and the spectral fine structure, respectively.

*韓國電子通信研究所
Electronics and Telecomm. Research Institute
論文番號 : 94147
接受日字 : 1994年 6月 3日

The spectral envelope is determined by the shape of vacal tract, while the spectral fine structure is caused by the periodic vibration of the vocal tract. In a speech coding system, a pitch predictor is used to remove long-term redundancies due to the periodicities of voiced part in speech signals. The pitch predictor has parameters such as a pitch period and pitch predictor coefficients. The pitch predictor should be updated because the actual pitch period and the amount of long-term correlation in the speech signal may vary in time.

There are two methods for adapting the pitch predictor : forward adaptation and backward adaptation. As in the formant predictor, the forward adaptation requires the transmission of side information to the receiver and long encoding delay due to buffering of input speech samples. For the low delay coder, the pitch predictor should be updated backwardly. The pitch period and filter coeffcients should be estimated using reconstructed signals which are available in the decoder. Recently the low delay vector excitation coding (LD-VXC) at 16 kbps which includes backwardly adapted pitch predictor was propoed by Cuperman [1]. The LD-VXC codec showed 2 dB improvement by adding backwardly adapted pitch in the speech coder with only formant predictor.

The backward adaptive pitch predictor is known to be very sensitive to channel errors. The transmission error effects are propagated very extensively because the pitch synthesis filter has an all-pole structure in which its impulse response becomes substantially long due to the distant sample prediction. In the LD-VXC, system performance was improved somewhat in the preesence of transmission error by adapting parallel adaptation between formant predictor and pitch prdictor, and including the inhibition control in unvocied/voiced speech. But the hybrid pitch predictor used in the LD-VXC did not show satisfactory performance in the high error rate channels. In this paper we propose a method to improve the robustness of pitch predictor in the noisy channel and the codec performance in the error free channel by tracking the change of pitch period within a block. The robust pitch predictor algorithm is operated in conjunction with a backwardly adapted tree coder at 9.6 kbps and TDHS(Time domain harmonic scaling)-tree coder at 6.4kbps [2].

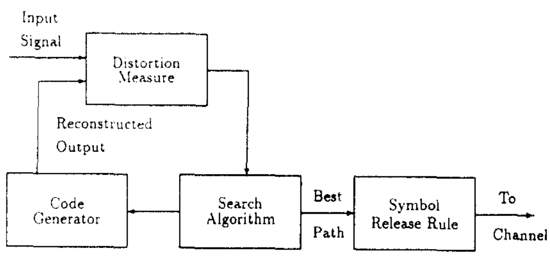

Fig. 1. Elements of a Tree Coder

II. TREE CODING OF SPEECH

Tree coding is a technique of multi-path searching used to improve the performance of a standard predictive coding approach by delaying the encoding decision for L samples which makes it possible to achieve higher quality coding [3]. A block diagram of a typical tree coder is shown in Fig.1 It consists of four functional elements: a code generator, a distortion measure, a search algorithm, and a symbol release rule. The code generator produces reconstructed output sequences according to all possible path maps. This reconstructed output sequence is transmitted to the distortion measure block to calculate the distortion between the source sequence and reconstructed candidate output sequence to the depth L. The calculated distortion is given to the search algorithm block to find the best path map with the smallest distortion. The best path is selected among the set of paths that is determined by a algorithm.

A. Tree Codes and Gain Adaptation

The excitation suquences are sequences of symbols taken from a tree structure. Various rate code trees can be constructed by choosing the number of branches per node and the number of symbols per branch. In our work, two code trees which have 2bits /sample and 1.5 bits /sample, respectively, will be considered. A 2 bits /sample vector code tree can be constructed that has 16 branches /node and 2 symbols /branch where the branch labels are populated with pairwise combinations of four-level minimum mean square error (MMSE) Gaussian quantizer outputs. For 1.5 bits /sample, a multitree code can be constructed by interleaving sequences fo cascaded 4-level and 2-level trees. These branches are populated with 2 bits and 1 bit MMSE quantizer.

The gain term should be adapted to the input signal because the speech signal is nonstationary and has relatively large input amplitude dynamic range. The gain adaptation method is dependant on the structure of the code tree. The hybrid adaptive quantizer[4].is used in 4-4 code tree and the hybrid adaptive quantizer and delta modulation algorithm are used in multitree code. The hybrid gain adaptation rule combining the instantaneous adaptation and syllabic dadptation is given by

$$\triangle(n+1)=M(n)\delta(n)\beta(n)\triangle^{\beta}(n) \tag{1}$$

with $\delta(n)=\alpha\delta(n-1)+(1-\alpha)\theta|u(n)|$ where $\alpha=0.9$ controls the effective memory of the estimator, θ is a constant, and $u(n)$ is the latest quantizer output. In the hybrid quantizer, $\beta=50/64$, $\gamma=13/64$, and $\theta=1.253$ are chosen.

B. Short-term Predictor

A recent modification [2] to the class of parameter estimation algorithm wherein the algorithm input sequence is shaped or smoothed is employed and critical to the performance obtained. It is important to choose the input to the adaptation in the pole-zero predictor for the robustness of tree coding to channel errors. In the pole-zero structure for a short-term predictor, the all-zero shaping filter is obtained by truncating the impulse response of the pole-zero transfer function. The all-zero shaping filter is chosen to satisfy,

$$\frac{1-B(z)}{1-A(z)} \simeq 1+D(z) \tag{2}$$

Then, the shaping flter coefficients, d_k are obtained by

$$d_k=\begin{cases}\sum_{j=1}^{N} a_j d_{k-j}+b_k & 1\leq k\leq M \\ \sum_{j=1}^{N} a_j d_{k-j} & M\leq k\leq p\end{cases} \tag{3}$$

where M is the order of the all-zero predictor. The filtered residual signal is given by

$$\tilde{e}_d(n)=u(n)+\sum_{K=1}^{P} d_k u(n-k) \tag{4}$$

Fig. 2 shows a configuration for the filtered residual signal as the input of the all-pole short-term predictor adaptation. The order of the shaping filter $1+D(z)$ is given by 8.

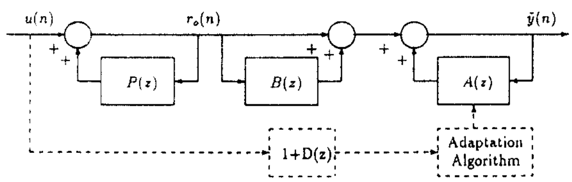

Fig. 2. Filtered Residual Driven Methods for All-Pole Predictor in Pole-Zero Short-term Predictor

Ⅲ. BACKWARD ADAPTED PITCH PREDICTION

The three-tap long-term(pitch) predictor has the transfer function,

$$P(z)=\beta_{-1}{}^{-(M-1)}+\beta_0 z^{-M1}+\beta_1{}^{-(M+1)} \tag{5}$$

where M_1 is the pitch period and β_{-1}, β_0, β_1 are long-term predictor coefficients. In a backward block adaptation for the pitch predictor, the pitch period and pitch predictor coefficients are estimated from a block of previously reconstructed pitch synthesizer outputs. The pitch period estimate M_1 is determined by searching the lag k, at time n, with which the normalized correlation function $\rho_k(n)$ is maximized. The pitch period M_1 is found in some bounded range. The normalized correlation function $\rho_k(n)$ is given by

$$\rho_k(n) = \frac{\Phi(0, k)}{\sqrt{\Phi_n(0,0)\Phi_n(k,k)}} \tag{6}$$

where the covariance function $\Phi_n(j,k)$ of the pitch synthesizer output is given by

$$\Phi_n(j,k) = \sum_{i=1}^{N_s} r_o(n-N_s+i-j)\, r_o(n-N_s+i-k) \tag{7}$$

Here, N_s is the number of samples in an analysis frame and $r_o(i)$ is the pitch synthesizer output.
After the pitch period M_1 is decided, pitch predictor coefficients are calculated by solving the Wiener-Hopf equation, so that

$$\begin{bmatrix} \Phi_n(M_1-1, M_1-1) & \Phi_n(M_1-1, M_1) & \Phi_n(M_1-1, M_1+1) \\ \Phi_n(M_1, M_1-1) & \Phi_n(M_1, M_1) & \Phi_n(M_1, M_1+1) \\ \Phi_n(M_1+1, M_1-1) & \Phi_n(M_1+1, M_1) & \Phi_n(M_1+1, M_1+1) \end{bmatrix}$$

$$\begin{bmatrix} \beta_{-1} \\ \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \Phi_n(0, M_1-1) \\ \Phi_n(0, M_1) \\ \Phi_n(0, M_1+1) \end{bmatrix}$$

A method in which the pitch predictor parameters are recursively updated at every sample in between the block adaptation was developed by Pettigrew and Cuperman [1]. This recursive backward adaptation consists of pitch period tracking and pitch predictor by examining the autocorrelation values of a present pitch synthesizer output and three samples that are pitch period lagged at every sample. The pitch predictor coefficients are updated by the gradient algorithm as

$$\beta_k(n) = \lambda\beta_k(n-1) + \frac{\mu_s}{\hat{\sigma}_{e_q}(n)\hat{\sigma}_{r_o}(n)}\, e_q(n)\, r_o(n-M_1+k)$$

$$k = -1, 0, 1 \tag{8}$$

where the leakage factor λ is given by 0.95 and the constant step size μ_s is given by 0.06
The variance of $\hat{e}_q(n)$ and $\hat{r}_o(n)$ are updated by

$$\hat{\delta}_x^2(n) = \lambda\hat{\delta}_x^2(n-1) + (1-\lambda)x^2(n) \tag{9}$$

IV. ROBUST ADAPTIVE PITCH PREDICTION

The use of a pitch predictor in linear predictive coding system is effcient way to represent the periodicity in the speech signal. The predictor gain depends on many factors which consist of the predictor order, input sampling frequency, the amount of periodicity in the input signal, and the updating algorithm of pitch predictor parameters. The prediction gain increases as sampling rate increases because a mismatch between real pitch period and its representation by integer multiples of the sampling interval is reduced. The order of the pitch predictor is typically 1 or 3. The use of multiple predictor cefficients provides an interpolation for pitch periodicities that are not an integer value and some additional prediction gain over a single tap pitch predictor.

An important fact the pitch predictor is that the all-pole structure of the pitch synthesizer is very sensitive to channel errors since its impulse response becomes substantially long due to the distant sample prediction. Therefore, we modify the input of the pitch predictor to be given by the interpolation of neighborhood samples. The smoother with 3 taps has a transfer function $S(z) = s_1 z^{-1} + s_0 z^0 + s_{-1} z^{1}$ and is placed in the location shown in Fig. 3. The adaptation scheme of the pitch predictor is illustrated in Fig. 4. The pitch predictor becomes more robust if the coefficients of the smoother are chosen to implement the function of a low pass filtering. Moreover, in order to track the pitch period change in a block, the coef ficients of the smoother can be variable according to the autocorrelation values of the present pitch synthesizer output and three lagged samples with pitch period. The coeffcients of the smoother are assigned as $s_1 > s_0 > s_{-1}$ if the following conditions are satisfied :

$$\hat{\rho}_{M_1+1}(n) > \hat{\rho}_{M_1}(n)$$

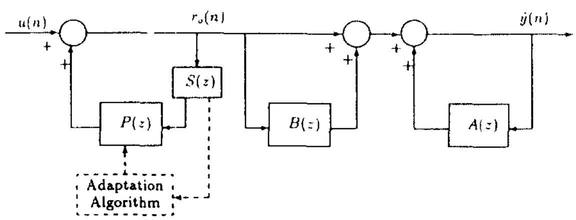

Fig. 3. Pitch Predictor with Smoother

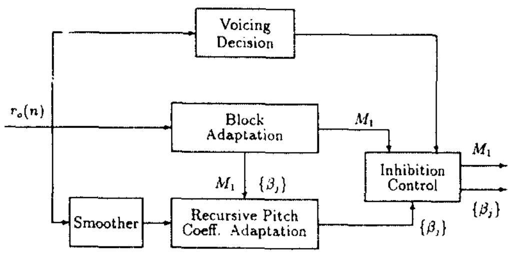

Fig. 4. Block Diagram for Pich Predictor Adaptation with a Smoother

$$\hat{\rho}_{M_1+1}(n) > \hat{\rho}_{M_1-1}(n)$$
$$\hat{\rho}_{M_1+1}(n) > \hat{\rho}_{min}$$

The autocorrelation function $\hat{\rho}_k(n)$ is estimated by the follwing recursion,

$$\hat{\rho}_k(n) = \lambda\hat{\rho}_k(n-1) + \frac{r_0(n)\, r_0(n-k)}{\hat{\sigma}^2_{r_0}(n)} \tag{10}$$

where the leakage factor λ is 0.95 and the variance of the pitch synthesizer output, $\hat{\sigma}^2_{r_0}(n)$ is updated by (9). The coefficient s_1 is weighted more than s_0 and s_{-1} because the pitch period of the input is considered to increase by one sample in this case. The coefficients of the smoother are assigned as $s_{-1} > s_0 > s_1$ if the following conditions are satisfied:

$$\hat{\rho}_{M_1-1}(n) > \hat{\rho}_{M_1}(n)$$
$$\hat{\rho}_{M_1-1}(n) > \hat{\rho}_{M_1+1}(n)$$
$$\hat{\rho}_{M_1-1}(n) > \hat{\rho}_{min}$$

The coefficients s_{-1} is weighted more than s_0 and s_1 because the pitch period of the input is considered to decrease by one sample. Otherwise, s_{-1}, s_0 and s_1 are the same as thst of the fixed smoother.

The coefficients of the fixed smoother are given by (0.25, 0.5, 0.25), and coefficients used in the variable smoother are decided either (0.53, 0.39, 0.08) or (0.08, 0.39, 0.53)depending on autocorrelation values of the pitch synthesizer output. The frequency response magnitude of the smoother is shown in Fig. 5. It is clear that smoother performs low pass filtering. This interpolated pitch synthesizer output is used as the input in the recursive pitch coefficient adaptation and the calculation of the pitch prediction value. The gradient recursive algorithm is written by

$$\beta_k(n) = \lambda\beta_k(n-1) + \frac{\mu_s}{\hat{\sigma}_{e_q}(n)\hat{\sigma}_{r_s}(n)}\, e_q(n)\, r_s(n-M_1+k)$$

$$k = -1,\ 0,\ 1 \tag{11}$$

where

$$r_s(n-M_1+k) = s_1 r_0(n-M_1+k-1) + s_0 r_0(n-M_1+k) + s_{-1} r_0(n-M_1+k+1) \tag{12}$$

The pitch synthesizer output is

$$r_0(n) = e_q(n) + \sum_{K=-1}^{1} \beta_k r_s(n-M_1+k) \tag{13}$$

$$e_q(n) + \sum_{K=-1}^{1} \beta'_k r_o(n-M_1+k) \tag{14}$$

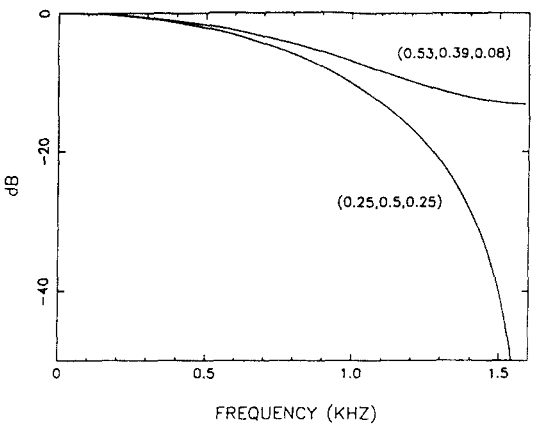

Fig. 5. Frequency Response of Smoother

where

$$\beta'_{-2} = \beta_{-1}s_1$$
$$\beta'_{-1} = \beta_{-1}s_0 + \beta_0 s_1$$
$$\beta'_0 = \beta_{-1}s_{-1} + \beta_0 s_0 + \beta_1 s_1$$
$$\beta'_1 = \beta_0 s_{-1} + \beta_1 s_0$$
$$\beta'_2 = \beta_1 s_{-1}$$

The equation (14) has the effect of increasing the number of pitch predictor taps to 5 from 3

V. PERFORMANCE OF BACKWARD ADAPTIVE PITCH PREDICTORS

The performance comparison of the pitch predictor with Cuperman's hybird adaptation and a hybrid adaptation with a variable smoother are presented in Table I and Table II for a TDHS-tree coder at rate of 6.4kbps and a tree coder at 9.6kbps, respectively. Here, Fig 6 shows the improvement by the inclusion of a smoother for the 6.4 kbps TDHS-tree coder.

The hybrid adaptation with a pitch tracker is very sensitive to channel errors and does not contribute largely to the performance of the tree coder. The use of the smoothed input in the pitch predictor gives a notable error performance increase about 4 dB at 10^{-3}BER

Table I. Performance Comparison of Hybrid Pitch Predictors in 6.4 Kbps TDHS-Tree Coder

Sent.	BER	SNR/SNRSEG (values in dB)			
		Cuperman's		Var. Smoother	
Fem.	0	21.17	19.71	21.61	20.21
	10^{-4}	15.70	16.26	19.36	19.01
	10^{-3}	8.17	9.18	13.35	13.86
	10^{-2}	3.18	3.01	4.43	4.70
Male	0	13.20	16.96	15.31	17.56
	10^{-4}	8.46	12.29	14.63	16.78
	10^{-3}	4.63	7.16	11.81	13.20
	10^{-2}	-0.26	1.79	2.79	3.86

Table II. Performance Comparison of Hybrid Pitch Predictors in 9.6 Kbps Tree Coder

Sent.	BER	SNR/SNRSEG (values in dB)			
		Cuperman's		Var. Smoother	
Fem.	0	18.19	17.29	18.13	17.48
	10^{-4}	16.77	16.30	17.28	16.93
	10^{-3}	11.23	10.91	13.10	12.84
	10^{-2}	4.91	4.45	6.19	6.30
Male	0	12.25	14.66	12.53	15.22
	10^{-4}	11.63	13.28	12.33	14.71
	10^{-3}	9.32	9.53	10.32	11.37
	10^{-2}	4.16	4.24	5.15	5.42

and about 2 dB at 10^{-3} BER over Cuperman's hybrid pitch predictor. The pitch predictor with a variable smoother provides an increase of about 0.6 dB in SNRSEG over the pitch predictor without a smoother in the noise free channe. The performance improvement is notable in the low pitched male speech. The pitch predictor with a variable smoother can track the pitch period and its representation by integer multiples of the sampling interval due to a sampling rate reduction. Fig. 7-Fig. 9 show narrowband spectrograms of original speech and two pitch predictor schemes at a BER of 10^{-3}. The pitch predictor without a smoother shows considerable distortion in har-

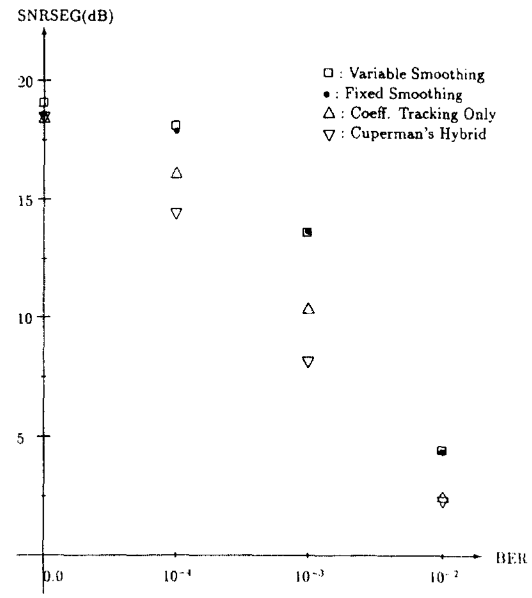

Fig. 6. Performance Comparison of Various Pitch Predictor Schemes for 6.4kbps

monic structure. However, the pitch predictor with a smoother maintains the periodic structure of speech well. In listening tests, the perceptual improvement is evident and it maintains good quality speech even at BER of 10^{-3}.

For the multitree coder at 9.6kbps, the performance improvement in the noise free channel is reduced a little because the pitch period mismatch is reduced by increasing the input sampling rate of tree coder. For 9.6 kbps multitree coder, the pitch predictor with a smoother provided about 0.35 dB gain in SNRSEG over the pitch predictor without a smoother. Moreover, the smoother inclusion in the pitch predictor gave us much robustness in the noisy channel.

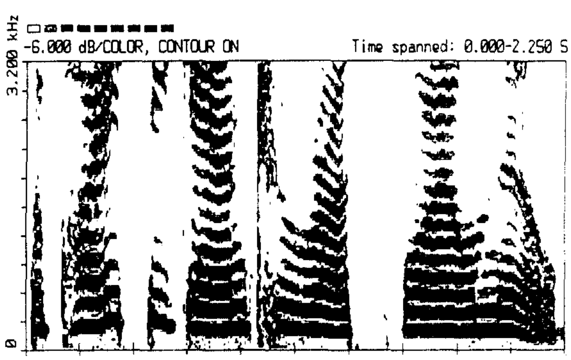

Fig. 7. Spectrogram of Original Speech

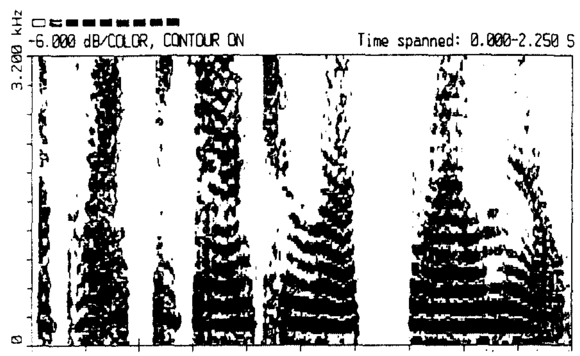

Fig. 8. Spectrogram of Reconstructed Speech in Noisy Channel for 6.4kbps TDHS-TREE Coder:Cuperman's pitch predictor, BER=10^{-3}

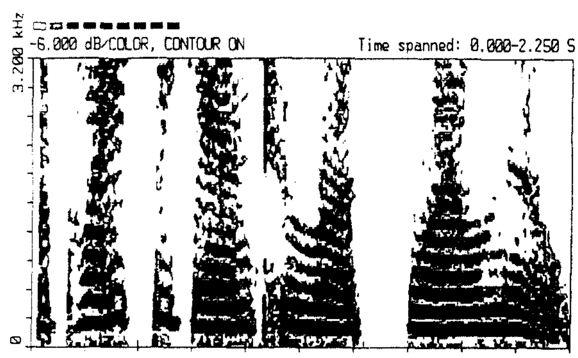

Fig. 9. Spectrogram of Reconstructed Speech in Noisy Channel for 6.4kbps TDHS-TREE Coder:With Variable Smoother, BER=10^{-3}

Ⅵ. CONCLUSIONS

At a low rate, speech quality depends primarily on the accurate representation of voiced speech. Maintaining the correct degree of periodicity in voiced speech is improtant to its perceptual quality. In order to improve the error performance and track the pitch period change of the input speech, it was proposed to smooth the input of the pitch predictor. The smoother with three taps can have fixed coefficients or variable codfficients depending on the estimated autocorrelation function of the output of the pitch synthesizer. The incusion of smoother in the pitch predictor not only increase the robustness of speech coder, but also improve the mismatch of pitxh period estimation. It gave a 0.6 dB gain in SNRSEG over the pitch predictor without a smoother in the noise free channel. The performance inprovement was notable in the low pitched male speech. In the noisy channel, it gained up to 4 dB in SNRSEG at BER of 10^{-3}, and 2.5 dB in SNRSEG at BER of 10^{-2}. In the subjective listening test, tree coding with robust pitch predictor contains little perceived distortion at the BER of 10^{-3}. At the BER of 10^{-2}, the reconstructed speech is intelligible and easily understandable.

Though a robust backward adpataive pitch predictor was applied in the tree coder, this pitch predictor can be applied in other backward adapted speech coders such as the low delay CELP (LD CELP)[5] which is recently adopted by CCITT as a 16 kbps standard and future low delay speech coders below 16 kbps.

REFERENCES

1. R. Pettigrew and V. Cuperman, "Backward pitch prediction for low delay speech coder", in Conf. Rec. IEEE Global Telecommun. Conf., pp. 1247-1252, 989.
2. I. Lee and J. D Gibson, "Tree coding combined with harmonic scaling of speech at 6.4kbps," in Proc. IEEE Conf. Acoust. Speech, Signal Processing , pp. 600-603, Apr, 1993.
3. J. D. Gibson and G. B. Haschke, "Adaptive code genetators for tree coding of speech," in Proc. IEEE Int. Conf. Commun., pp. 1142-1146, June 1987.
4. I. Lee, "Robust tree coding combined with harmonic scaling of speech,". Ph.D. dissertation. Dept. of Elec. Eng., Texas A&M University, College Station, TX, Dec. 1992.
5. J. H. Chen, "A robust low-delay CELP speech coder at 16 kbit/s," in Conf. Rec. IEEE Global Telecommun. Conf., pp. 1237-1241, 1989.

李 寅 誠(Insung Lee) 정회원
1960년 6월 8일생
1983년 2월 : 연세대학교 전자공학과 졸업(공학사)
1985년 2월 : 연세대학교 전자공학과 대학원 졸업(공학석사)
1992년 12월 : 미국 Texas A&M Univ. 전기공학과 졸업(Ph.D.)
1986년 5월 ~ 1987년 7월 : 한국통신 연구개발단 전임연구원
1993년 2월 ~ 현재 : 한국전자통신연구소 신호처리연구실 선임연구원
※주관심분야 : 음성, 영상데이타 압축, 적응필터, 이동통신