Robust Speech Decoding Using Channel-Adaptive Parameter Estimation.

*Yun-Keun Lee, **Hwang-Soo Lee

Abstract

In digital mobile communication system, the transmission errors affect the quality of output speech seriously. There are many error concealment techniques using a posteriori probability which provides information about any transmitted parameter. They need knowledge about channel transition probability as well as the 1st order Markov transition probability of codec parameters for estimation of transmitted parameters. However, in applications of mobile communication systems, the channel transition probability varies depending on nonstationary channel characteristics. The mismatch of designed channel transition probability of the estimator to actual channel transition probability degrades the performance of the estimator.

In this paper, we proposed a new parameter estimator which adapts to the channel characteristics using short time average of maximum a posteriori probabilities(MAPs). The proposed scheme, when applied to the LSP parameter estimation, performed better than the conventional estimator which do not adapt to the channel characteristics.

I. Introduction

In digital mobile communication system, it is required to reduce transmission errors which affect the quality of output speech seriously. There are some earlier publications about error concealment or enhancing robustness against transmission errors. Many of them utilize the residual redundancy of the source coder [1,2] and estimate the transmitted bits or parameters using a posteriori probability [3,4]. These methods need knowledge about channel transition probability for computation of a posteriori probability. However, in applications of mobile communication systems, the channel transition probability varies depending on nonstationary channel characteristics. The mismatch of designed channel transition probability of the estimator to actual channel transition probability degrades the performance of the estimator.

In this paper, we proposed a channel-adaptive parameter estimation(CAPE) method which matches the estimator to the nonstationary channel characteristics using short time average of MAPs. In the next section, we describe outline of the proposed scheme. In sections 3 and 4, channel model and parameter estimation methods are presented. In section 5, we describe the principle of the proposed scheme and simulation results are discussed in section 6.

* Information Technology Lab., LG Corporate Institute of Techn ology

** Department of Electrical Engineering, KAIST

Manuscript Received : October 21, 1998.

II. Outline of the scheme

The outline of the proposed speech decoding scheme is shown in Fig. 1. This scheme shows the procedure of data transmission. At the encoder, the source parameter v is quantized and converted into the bit pattern $x^{(i)}$. The bit pattern is transmitted to the receiver through the channel. At the decoder, reconstructed parameter $\hat{v}$ is estimated by CAPE algorithm using the received parameter $\hat{x}$. The following notation is used in the explanation of the proposed scheme.

v : Source parameter

v_q : Quantized source parameter

$\hat{v}$: Reconstructed source parameter

M : Number of bits for a source parameter

$x^{(i)}$: i-th bit pattern before transmission, $0 \le i \le 2^M - 1$

$\hat{x}$: Received bit pattern

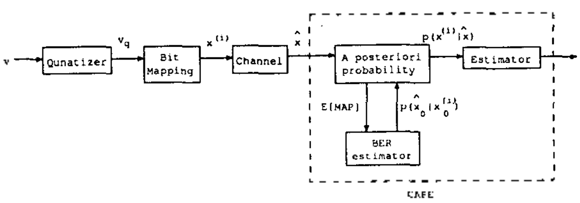

Figure 1. Outline of the scheme.

III. Channel model

The probability to receive bit $\hat{x}(m)$ if bit $x^{(i)}(m)$ was transmitted can be written as

$$p(\hat{x}(m)|x^{(i)}(m)) = \begin{cases} 1-p_e & \text{if } \hat{x}(m) = x^{(i)}(m) \\ p_e & \text{if } \hat{x}(m) \neq x^{(i)}(m) \end{cases} \quad (1)$$

where p_e denotes the channel bit error rate(BER). If a data x consists of M bits and the channel is memoryless, the transition probability is

$$p(\hat{x} \mid x^{(i)}) = \prod_{m=0}^{M-1} p(\hat{x}(m)|x^{(i)}(m)). \quad (2)$$

IV. Parameter estimation

Generally, the speech coding schemes aim at minimizing the redundancy of source parameters. However, there is residual interframe correlations in most speech coding schemes. Therefore, we can use information of already received parameters on estimating the parameters of current frame. The source parameter can be estimated at the receiver using a posteriori probability $p(x^{(i)}_0 \mid \hat{x}_0, \hat{x}_{-1}, \hat{x}_{-2}, \cdots)$, where $x^{(i)}_0$ is the transmitted parameter of current frame and $\hat{x}_i$ is the received parameter of i-th frame. With an assumption that the sequence of source parameters is 1st order Markov process, the a posteriori probability can be given in a recursive form [4] as

$$\begin{aligned} p(x^{(i)}_0 \mid \hat{x}_0, \hat{x}_{-1}, \hat{x}_{-2}, \cdots) = C \cdot p(\hat{x}_0 \mid x^{(i)}_0) \cdot \\ \sum_{j=0}^{2^M-1} p(x^{(i)}_0 \mid x^{(j)}_{-1}) \cdot \\ p(x^{(j)}_{-1} \mid \hat{x}_{-1}, \hat{x}_{-2}, \hat{x}_{-3}, \cdots) \end{aligned} \quad (3)$$

The term $p(x^{(i)}_0 | x^{(j)}_{-1})$ is the 1st order Markov transition probability and $p(\hat{x}_0 | x^{(i)}_0)$ is a channel transition probability.

There are two major principles of estimation such as maximum a posteriori probability (MAP) estimation and minimum mean square error (MMSE) estimation [5]. The MAP estimator delivers an estimated parameter value according to

$$\hat{v}_{MAP} = v_q^{(\mu)} \text{ with } \mu = \underset{i}{\operatorname{argmax}}\, p(x^{(i)} \mid \hat{x}_0, \hat{x}_{-1}, \hat{x}_{-2}, \cdots). \quad (4)$$

where $v_q^{(\mu)}$ is the quantized parameter value of index μ. The MMSE estimator delivers the optimum in a minimum mean square error sense by the expectation value

$$\hat{v}_{MMSE} = \sum_{i=0}^{2^M-1} v_q^{(i)} \cdot p(x^{(i)} \mid \hat{x}_0, \hat{x}_{-1}, \hat{x}_{-2}, \cdots). \quad (5)$$

In experiments of LSP parameter estimation, the MMSE estimation performs better than the MAP estimation.

V. Channel-Adaptive Parameter Estimation (CAPE)

Parameter estimation using a posteriori probability requires the information about BER of the channel to compute the transition probability $p(\hat{x}_0 | x_0^{(i)})$. However, in applications of mobile communication systems, the BER varies depending on nonstationary channel characteristics. The mismatch of designed channel BER of the estimator(BERE) to actual channel BER-(BERC) causes performance degradation in parameter estimation. We simulate the LSP parameter estimation on channels of various BERCs using MMSE estimators of various BEREs. Fig. 4 shows spectral distances(SDs) [6] between spectrum of transmitted LSP parameters and estimated LSP parameters at the receiver. Since each estimator excels when the BERE equals to the BERC, it is required to adapt the BERE to the channel characteristics for improving the performance of the estimator.

The a posteriori probability $p(x_0^{(i)} | \hat{x}_0, \hat{x}_{-1}, \hat{x}_{-2}, \cdots)$ tend to decrease as the channel error increases. We computed the MAPs, $\max p(x_0^{(i)} | \hat{x}_0, \hat{x}_{-1}, \hat{x}_{-2}, \cdots)$ at various BERE and BERC by computer simulations on estimation of LSP parameters. We estimate each LSP parameter using already received LSP parameters of the same order by equation (3) and (5). We also computed the values of MAPs every frame. We repeated this experiment varying BERE and BERC. We found out that the expectation value of the MAPs, E[MAP] has some correlations with the BERE and the BERC. The E[MAP] increases as the BERE and the BERC decreases as shown in Fig. 2. Fig. 2 demonstrates that the E[MAP] can be expressed with second order polynomials of the BERC, given fixed BERE, and each coefficient of polynomials also can be expressed with second order polynomials of the BERE. The expression of the E[MAP] in terms of the BERC and the BERE is

$$E[MAP] = a(BERE) \cdot BERC^2 + b(BERE) \cdot BERC + c(BERE)$$

$$\begin{aligned} a(BERE) &= a_1 \cdot BERE^2 + a_2 \cdot BERE + a_3 \\ b(BERE) &= b_1 \cdot BERE^2 + b_2 \cdot BERE + b_3 \\ c(BERE) &= c_1 \cdot BERE^2 + c_2 \cdot BERE + c_3 \end{aligned} \qquad (6)$$

The coefficients of polynomials obtained by regression analysis are shown in Table 1. The current value of BERC can be estimated by (6) if the E[MAP] and BERE are given. Since the channel characteristics are time-varying, we use short time average of MAPs instead of the E[MAP]. The estimator adapts to the channel characteristics by updating the BERE every frame using the estimated BERC.

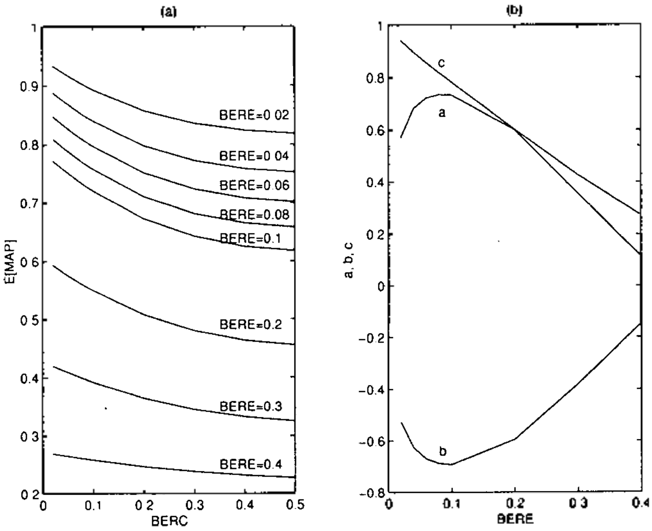

Figure 2. (a) E[MAP] vs. BERC at various BEREs
(b) Coefficients of Polynomials vs. BERE.

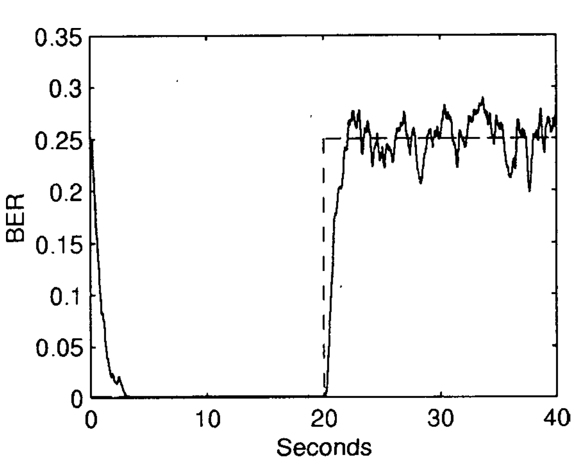

Figure 3. Adaptation of BERE (dashed line: BERC, soline: BERE).

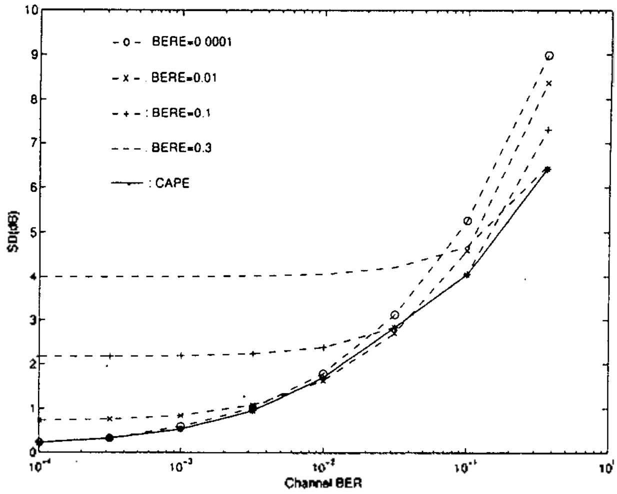

Figure 4. Performances of the estimators over various channel conditions.

Table 1. Coefficients of polynomials.

	BERE ≤ 0.1	BERE > 0.1
a_1	−45.3669	- 2.6645
a_2	7.3528	- 0.7659
a_3	0.4474	0.8433
b_1	38.8125	3.3973
b_2	- 6.173	0.1493
b_3	- 0.4149	- 0.7462
c_1	2.7450	0.6366
c_2	- 2.2892	- 2.0312
c_3	0.9858	0.9816

VI. Experiments and results

We simulated LSP parameter estimations by MMSE criterion in order to measure the performance of the CAPE algorithm. The speech database is LPC analyzed and ten LSP parameters were obtained at every frame of 20 ms. The speech database consists of about 8 minutes of sentences. Each LSP coefficient is nonlinearly quantized with 3 or 4 bits [7]. We compute the 1st order Markov transition probability $p(x^{(i)}_0 | x^{(i)}_{-1})$ using the sequence of encoded LSP vectors.

The quantized LSP parameters were transmitted to the decoder over noisy channel. We simulated the noisy channel by inserting random bit error of given BERC into the encoded bit stream. At the decoder, the transmitted LSP parameters were estimated using the 1st order Markov transition probability and BERE which adapts dynamicaly to the BERC by CAPE algorithm.

We observe the adaptation process of the estimator when the BERE mismatches to the BERC. The BERE is updated every frame to track the BERC by

$$BERE_{curr} = BERE_{prev} + \alpha(BERC_{curr} - BERE_{prev}) \quad (7)$$

where α is convergence factor. We can compute the $BERC_{curr}$ by equation (6) using the $BERE_{prev}$ and the short time average of MAPs. The duration of average of MAPs and the convergence factor were set to adequate value of 10 frames and 0.1, respectively by computer simulation. An example of trajectory of BERE is shown in Fig. 3. Only a few seconds are sufficient to match the BERE to the target BERC. We compared the performance of the CAPE with conventional estimator which have fixed BERE. The performances of the estimators were measured by SDs between transmitted and reconstructed spectra. The speech database consists of about 30 seconds of sentences which were pronounced by 2 male and 2 female speakers. Fig. 4 demonstrates that the performances of CAPE are similar to the best performances of the conventional estimators at every channel BERs.

VII. Conclusions

In this paper we discussed the parameter estimation schemes using a posteriori probability. We simulated LSP parameter estimations by MMSE criterion and demonstrated that the mismatch of the designed channel BER of the estimator to the actual channel BER degrades the performance of the estimator.

The parameter estimation method CAPE was proposed, which matches the estimator to the nonstationary channel characteristics using short time average of MAPs. We derived equations of expectation value of MAPs in terms of the designed channel BER of the estimator and the actual channel BER. The channel BER of current frame is estimated using short time average of the MAPs. The estimator adapts to the channel characteristics in a few seconds using the estimated channel BER. The proposed scheme when applied to the LSP parameter estimation performed better than the conventional estimator which do not adapt to the channel characteristics.

References

1. K. Sayood and J. C. Borkenhagen, "Use of residual redundancy in the design of joint source/channel coders," *IEEE Tran. on Comm.*, vol. COM-39, no. 6, pp. 838-846, June, 1991.
2. S. Feldes, "Enhancing robustness of coded LPC-spectra to channel errors by use of residual redundancy," *Proc. of EUROSPEECH*, pp. 1147-1150, 1993.
3. J. Hagenaurer, "Source-controlled channel decoding," *IEEE Tran. Comm.*, vol. COM-43, no. 9, pp. 2449-2457, Sep., 1995.
4. T. Fingscheidt and P. Vary, "Robust speech decoding: a universal approach to bit error concealment," *Proc. of Int. Conf. on Acoustics, Speech, and Signal processing*, pp. 1667-1670, 1997.
5. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw Hill, New York, 1965.
6. K.K.Paliwal and B.S.Atal, "Efficient Vector Quantization of LPC parameters at 24 bits/frame," *IEEE Tran. on Speech and Audio Processing*, vol. 1, No. 1, pp.3-14, Jan. 1993.
7. A. M. Kondoz, *Digital Speech*, John Wiley & Sons, England, 1994.

▲Yun Keun Lee

Yun Keun,Lee was born in 1964. He received the B. S. degree in electrical engineering from Seoul National University in 1986, the M. S. degree and Ph. D degree in electrical engineering from Korea Advanced Institute of Technology in 1988 and 1998, respectively. He has been with LG Corporate Institute of Technology since 1988, where he is currently a senior engineer of Information Technology laboratory. His research interest includes speech signal processing, digital communication and application of digital signal processing techoloy.

▲Hwang Soo Lee

The Journal of the Acoustical Society of Korea (vol. 15, No. 2E, 1996)