Adaptive Encoding of Fixed Codebook in CELP Coders

*Hong Kook Kim

Abstract

In this paper, we propose an adaptive encoding method of fixed codebook in CELP coders and implement an adaptive fixed code exited linear prediction (AF-CELP) speech coder. AF-CELP exploits the fact that the fixed codebook contribution to speech signal is also periodic like the adaptive codebook (or pitch filter) contribution. By modeling the fixed code book with the pitch lag and the gain from the adaptive codebook, AF-CELP can be implemented at low bit rates as well as low complexity. Listening tests show that a 6.4 kbit/s AF-CELP has a comparable quality to the 8 kbit/s CS-ACELP in background noise conditions.

Ⅰ. Introduction

The pitch filter or the adaptive codebook in code-excited linear prediction (CELP) coders is employed to remove the periodicity of speech signal and the resultant signal is further modeled by a fixed coedvector. Accurate design of the structure and codewords of fixed codebook is important to dedcoded speech quality. To efficiently represent the excitation signal in CELP with low complexity, lots of schemes have been proposed such as a sparse codebook, an algebraic structured codebook, and so on [1]. Generally, the number of bits assigned to the fixed codebook reaches about 40 percent of total bits of a CELP coder. It makes CELP coders being hard to be realized in low bit rate. For example, the 8 kbits/s conjugate-structure algebraic CELP (CS-ACELP) [2] assigns 34 bits every 10 ms frame to the quantization of the algebraic codebook shape, where 34 bits corresponds to 42.5 percent of all bits. When we desire to obtain a low-bit-rate coder having the same structure to the above CS-ACELP, lowering the bits to represent the excitation signal causes the coder performance to degrade rapidly.

As a remedy to this problem, we have proposed the renewal excitation codebook approach to represent the excitation signal [3]. We first observed that the target residual used in fixed codebook search is periodic to some extent and correlated with the adaptive codebook excitation of speech signal. Therefore, the renewal excitation signal could be generated from the adaptive codebook. In this paper, we will propose an adaptive fixed codebook in CELP by extending the concept of the renewal excitation. Speech signal is medeled by filtering the combination of the adaptive codebook and the fixed codebook. Conventionally, the two codebooks are used to update the adaptive codebook memory. In the proposed approach, we have the adaptive fixed codebook memory designed separately from the adaptive codebook memory. Similar to updating the adaptive codebook, the adaptive fixed codebook is also updated with only the fixed codebook excitation.

*Digital Communication Lab., Samsung Advanced Institute of Technology

Manuscript Received: November 1, 1997.

Ⅱ. Adaptive Fixed Codebook

In order to observe the periodicity of the fixed codebook excitation, we obtained the waveforms as shown in Fig. 1. The target signal for the fixed codebook search has high amplitude near the pitch onset of the residual or adaptive codebook signal. Also, the pulses of algebraic excitation signal, which is resulted from the codebook search of the CS-ACELP, are mainly displaced at the regions corresponding to the target signal of high amplitude. From this, we say that the proposition that the fixed codebook excitation is periodic to some extent is reasonable.

Fig. 2 shows the block diagram for the proposed CELP coding algorithm. The excitation signal, $r(n)$, is represented as

$$r(n)=r_a'(n)+c_a'(n), \tag{1}$$

where $r_a'(n)$ and $c_a'(n)$ are the adaptive codebook excitation and the fixed codebook excitation, respectively. $c_a'(n)$ is produced from either the conventional fixed codebook or the adaptive fixed codebook. An adaptive fixed codevector without being multiplied by the fixed codebook

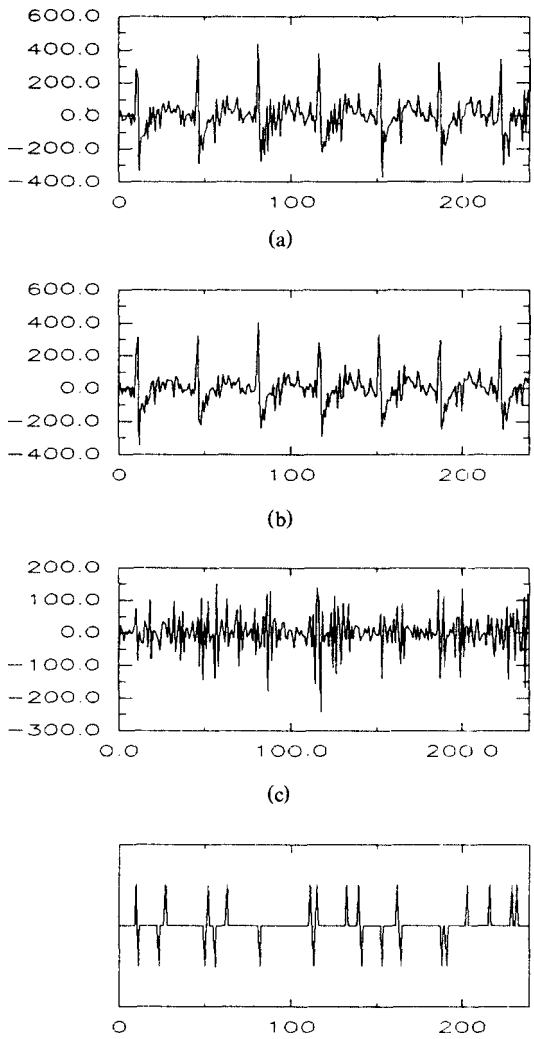


Figure 1. Waveforms: (a) the residual signal of a speech segment, (b) the adaptive codebook signal of the 8 kbit/s CS-ACELP, (c) the target residual which is the difference between (a) and (b), and (d) the algebraic codebook signal of the 8 kbit/s CS-ACELP

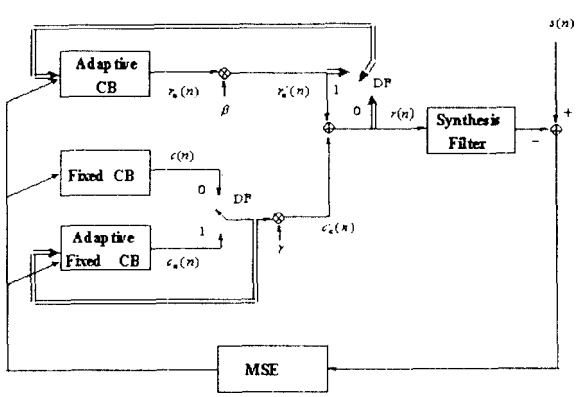


Figure 2. A block diagram of the proposed CELP coding algorithm

gain, γ, can be retrieved as follows:

$$c_a(n) = \beta_c c_a(n - T_c), \text{ for } 0 \le n < N \text{ and } T_c \ge N, \tag{2}$$

and for $T_c < N$

$$c_a(n) = \begin{cases} \beta_c c_a(n - T_c), & 0 \le n < T_c \\ \beta_c^2 c_a(n - 2T_c), & T_c \le n < N, \end{cases} \tag{3}$$

where N is the subframe size of a CELP coder, and T_c and β_c are the estimate of the periodicity and the gain of the adaptive fixed codevector, respectively.

For the adaptive fixed codebook (AFC) modeling, we should determine γ as well as T_c and β_c. It is strongly recommended that T_c and β_c are replaced with T_a and β_a corresponding to the pitch lag and the gain of adaptive codebook, respectively, in order to enable coders operate at low bit rates. From now on, we will denote $T_c = T_a = T$ and $\beta_c = \beta_a = \beta$ for the sake of simplicity. Compared to the pitch synchronous excitation modeling such as in the pitch synchronous innovation CELP (PSI-CELP) [4], there is no need to find the fixed codevector index in the proposed AFC modeling because the adaptive fixed codevector is selected from the AFC which only contains the history of past fixed codebook excitation signal. On the other hand, in the PSI-CELP, the fixed codevector is constructed by periodically repeating a codevector selected from the stored fixed codebook.

In order to obtain the AFC excitation parameters, we propose three search methods: sequential, joint, and hybrid optimum search. In the sequential optimum AFC search, the conventional adaptive codebook search is performed beforehand, and thus the pitch lag and the adaptive codebook gain are obtained. We assign T and β in (2) or (3) as the pitch lag and the gain of the adaptive codebook. The resultant overall error can be given by

$$\varepsilon(\gamma) = \sum_{n=0}^{N-1} (s(n) - s_0(n) - h(n) * r_a'(n) - \gamma c_a(n))^2, \tag{4}$$

where $s(n)$ is the original speech signal and $s_0(n)$ is the zero input response of the synthesis filter whose impulse response is $h(n)$. $r_a'(n)$ is previously known from the result of the adaptive codebook search and $c_a(n)$ is also obtained from (2) or (3) by substituting T and β. The fixed codebook gain minimizing (4) can be obtained as

$$\gamma = \frac{\sum_{n=0}^{N-1} \bar{s}(n) c_a(n)}{\sum_{n=0}^{N-1} c_a^2(n)}, \tag{5}$$

where $\bar{s}(n) = s(n) - s_0(n) - h(n) * r_a'(n)$. This AFC search

is so simple and has low complexity.

In the joint optimum search of AFC, T, β and γ are jointly optimized in analysis-by-synthesis loop. When $T \geq N$, the overall error energy for the joint search is given by

$$\varepsilon(T, \beta, \gamma) = \sum_{n=0}^{N-1} (\hat{s}(n) - \beta h_r(n-T) - \gamma\beta h_c(n-T))^2, \quad (6)$$

where $\hat{s}(n) = s(n) - s_0(n)$, $h_r(n-T) = h(n) * r_a(n-T)$ and $h_c(n-T) = h(n) * c_a(n-T)$. By differentiating (6) with respect to β and γ, respectively, we can obtain the optimum β and γ as

$$\beta = \frac{H_r \tilde{H}_c - H_c H_{rc}}{\tilde{H}_c \tilde{H}_r - H_{rc}^2}, \quad (7)$$

$$\gamma = \frac{H_c \tilde{H}_r - H_r H_{rc}}{H_r \tilde{H}_c - H_c H_{rc}}, \quad (8)$$

where $H_r = \sum_{n=0}^{N-1} \hat{s}(n) h_r(n-T)$, $H_c = \sum_{n=0}^{N-1} \hat{s}(n) h_c(n-T)$, $\tilde{H}_r = \sum_{n=0}^{N-1} h_r^2(n-T)$, $\tilde{H}_c = \sum_{n=0}^{N-1} h_c^2(n-T)$, and $H_{rc} = \sum_{n=0}^{N-1} h_r(n-T) h_c(n-T)$. For a given pitch lag T, we first compute β and γ by using the above equation. And then, they are substituted back into (6) to compute the squared error. We finally choose the optimal T, β, and γ which minimizes (6).

On the other hand, when the pitch lag is smaller than the subframe length, the overall squared error is also given by

$$\begin{aligned}\varepsilon(T, \beta, \gamma) = &\sum_{n=0}^{N-1} \hat{s}^2(n) - 2\beta \sum_{n=0}^{N-1} \hat{s}(n) h(n, T) \\ &+ \beta^2 \sum_{n=0}^{N-1} h^2(n, T) - 2\beta^2 \sum_{n=T}^{N-1} \hat{s}(n) h(n, 2T) \\ &- 2\beta^3 \sum_{n=T}^{N-1} h(n, T) h(n, 2T) + \beta^4 \sum_{n=T}^{N-1} h^2(n, 2T),\end{aligned} \quad (9)$$

where $h(n, t) = h_r(n, t) + \gamma h_c(n, t)$, for $t = T$ or $2T$, with

$$h_r(n, T) = \sum_{k=0}^{T-1} r_a(k-T) h(n-k),$$

$$h_r(n, 2T) = \sum_{k=T}^{N-1} r_a(k-2T) h(n-k),$$

and

$$h_c(n, T) = \sum_{k=0}^{T-1} c_a(k-T) h(n-k),$$

$$h_c(n, 2T) = \sum_{k=T}^{N-1} c_a(k-2T) h(n-k).$$

By setting the derivatives $\frac{\partial \varepsilon}{\partial \beta}$ and $\frac{\partial \varepsilon}{\partial \gamma}$ to be zero, we can obtain the closed form solutions of β and γ. However, the solution of β requires a root-finding procedure of the seventh order polynomal. This causes high computational burden. Therefore, we incorporate the quantization of β and γ into the joint optimum search procedure. Each quantized values of β and γ are substituted into (9) and the squared errors are computed over all pitch lags. Consequently, both T and the quantized β and γ are determined on a basis of minimum mean squared error criterion. This procedure can also be applied to the seach when $T \geq N$.

In practice, a hybrid search is preferred than the joint optimum search in a view of computational complexity. We first determine the optimal pitch lag by assuming the AFC contribution to be zero. In other words, the conventional adaptive codebook search is first performed. (6) and (9) are not any more function of T. When $T \geq N$, we can simply compute β and γ by using (7) and (8). Similary, for $T < N$, T is substituted into (9) and β and γ are obtained by solving the higher-order polynomials. In this hybrid search, we incorporate the quantization of β and γ into the equations of (6) and (9). We globally search codebook indices for β and γ by sequentially taking one of codewords from the quantization codebooks of β and γ and finding each codeword of β and γ minimizing the squared error. This procedure can significantly reduce the complexity of the joint optimum search.

Ⅲ. Implementation and Performance Evaluation

We designed a speech coder operating at 6.4 kbit/s by adopting the adaptive fixed codebook concept. The coder structure is based on the CS-ACELP operating at 8 kbit/s [2]. A 10 ms speech frame is divided into two subframes. The excitation signal for one of the two subframes is modeled by the proposed adaptive fixed codebook while the excitation for the other subframs is represented by using the conventional algebraic codebook.

Table 1 shows the bit allocation of the 6.4 kbit/s AF-CELP is the number of bits assigned to the fixed codebook shape and gain. There is no bit assigned to the fixed codebook in the AF-CELP. We first decide which subframe is adequate for the AFC modeling and assign 1 bit to this indicator which is called dynamic flag as shown in the fourth parameter of Table 1. For the selected subframe, the AFC method is applied to the fixed codebook modeling. On the contrary, the fixed codebook excitation of the other subframe is represented as the algebraic codebook. The adaptive codebook is updated differently ac-

cording to the value of the dynamic flag (DF). when the AFC search is employed (in the case of DF = 1 of Fig. 2), the adaptive codebook memory is updated without the fixed codebook excitation. On the other hand, the adaptive codebook memory is updated conventionally when DF = 0. As a search method, we employed the sequential AFC search method. In the CS-ACELP, the algebraic codebook search is a major portion of computational burden. We could reduce the search complexity by halving the number of algebraic codebook search and increasing a few computations in finding the AFC gain of (5).

Table 1. Bit allocation of the 6.4 kbit/s AF-CELP speech codei

Parameters	Assigned Bits		
	Subframe 1	Subframe 2	Frame
LSP	18		18
Pitch	8	5	13
Parity	1	–	1
Dynamic Flag	1		1
Fixed Codebook Shape	13(–)	–(13)	13
Fixed Codebook Sign	4(–)	–(4)	4
Gain	7	7	14
Total	64		

We have done subjective qualification tests for the AF-CELP. A degradation category rating (DCR) test over the 8 kbit/s CS-ACELP is carried out. The CS-ACELP is known to offer the toll-quality in various input conditions [5]. In DCR [6], the decoded speech processed by the CS-ACELP is presented before that processed by the AF-CELP. Ten listeners participated in the experiments and each listener judged the quality degradation of speech to be evaluated with regard to the preceding reference speech. Each degradation judgement is done on a 5-point degraded mean opinion score (DMOS). In order words, 5 point means the degradation with regard to the preceding speech is inaudible and 4 point degradation is audible but not annoying. Also, 3, 2, and 1 point are rated by the listeners when degradation is slightly annoying, annoying, and very annoying, respectively. Finally the DMOS of each presentation is collected and averaged over all listeners for each condition. We prepared speech data which consist of four sentences spoken by four Korean talkers (two males and two femals), where each sentence is of 8 sec long. Speeches are sampled with the rate of 16 kHz, and then filtered by the modified intermediate response system (IRS) filter followed by the automatic level adjustment [7]. To process them by the coders, they are downsampled from 16 kHz to 8 kHz.

First of all, the coder performance was tested under noise-free conditions, which include error-free, asynchronous two stage tandem capability, and random missing frames of 3%. As shown in Table 2, the AF-CELP shows the slightly degraded performance to the CS-ACELP under clean environment at which input level is set to −26 dBovl. However, the quality of the AF-CELP in frame erasure and tandeming conditions degrades annoyingly compared to that of the CS-ACELP.

Table 2. DCR test results of the proposed coder compared to the 8 kbit/s CS-ACELP under noise-free conditions

Conditions	DMOS		
	Male	Female	Avg.
Clean	4.40	4.05	4.23
2 Tandem	3.10	2.85	2.98
Frame Erasure 3%	3.60	3.55	3.73

Next, we tested the coder performance under three types of noise such as office babble at 30 dB signal-to-noise ratio (SNR), car noise at 15 dB SNR, and the interfering talker at 15 dB SNR. As shown in Table 3, the performance of the AF-CELP in babble and interfering talker noise conditions provides slightly audible degradation to that of the CS-ACELP. Under car noise condition, the AF-CELP smoothes noise signal as if it performs a kind of noise suppression. For this reason, most of listeners felt the difference of decoded speech qaulity from the AF-CELP, compared to that from the CS-ACELP.

Table 3. DCR test results of the proposed coder compared to the 8 kbit/s CS-ACELP under backgroud noise conditions

Conditions		DMOS		
	SNR	Male	Female	Avg.
Babble Noise	30 dB	4.40	4.20	4.30
Car Noise	15 dB	3.30	3.65	3.48
Interfering Talker	15 dB	4.50	4.40	4.45

In addition to the DCR test, we performed the absolute category rating (ACR) based on the 5-point MOS scale [6]. Table 4 and 5 show the MOS scores of the CS-ACELP and the AF-CELP under the same conditions

corresponding to Table 2 and 3, respectively. The MOS scores of the AF-CELP reach about 90% of those of the CS-ACELP for the test conditions of clean, babble and interfering talker. On the contrary, we could achieve the superior performance of 10% with the car noise condition while the degradation was observed for the tandeming condition. Fig. 3 shows the relation between the DMOS score and the ratio of the MOS score of the AF-CELP to that of the CS-ACELP for each test condition. From the figure, we could conclude that the two tests of the DCR and the ACR are consistent except the reference coder of lower quality.

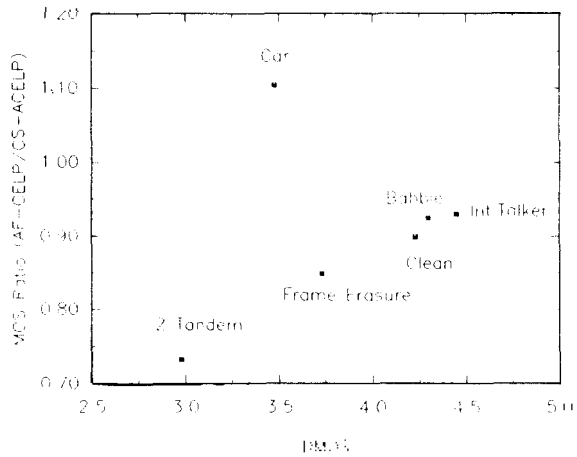


Figure 3. The ratios of the MOS score fo the AF-CELP to that of the CS-ACELP against the DMOS scores of the AF-CELP

Table 4. ACR test results of the CS-ACELP and the proposed coder under noise-free conditions

Conditions	CS-ACELP			AF-CELP		
	Male	Female	Avg.	Male	Female	Avg.
Clean	4.15	4.15	4.15	3.75	3.70	3.73
2 Tandem	3.15	3.30	3.275	2.35	2.55	2.40
Frame Erasure 3%	3.35	3.00	3.18	2.60	2.80	2.70

Table 5. ACR test results of the CS-ACELP and the proposed coder under background noise conditions

Conditions		CS-ACELP			AF-CELP		
	SNR	Male	Female	Avg.	Male	Female	Avg.
Babble Noise	30 dB	3.90	4.05	3.98	3.55	3.80	3.68
Car Noise	15 dB	2.80	2.95	2.88	3.00	3.35	3.18
Interfering Talker	15 dB	3.50	3.65	3.58	2.95	3.70	3.33

Finally, we designed a 7.2 kbit/s variable rate speech coder by switching speech coders between the 8 kbit/s CS-ACELP and the 6.4 kbit/s AF-CELP every 10 ms. An informal listening test shows that listeners could not differentiate the quality of the coder from that of the CS-ACELP.

We implemented the CS-ACELP and the 6.4 kbit/s AF-CELP on a Texas Instrument TMS320C5X fixed-point general purpose digital signal processor which runs at 40 MIPS. Table 6 shows the processor load comparison of the two coders. In our implementation of the CS-ACELP, the complexity of the algebraic codebook search reaches 5~10 MIPS. We can reduce the complexity by half because the AF-CELP searches the algebraic codebook once every frame. Of course, the authors [8] reported that algebraic codebook search requires 8.4 MIPS in worst case. We expect that about 50 persent reduction can be achieved in this case. On the other hand, about 10 percent increases of the program memory and the working memory are needed, respectively.

Table 6. Processor load of the CS-ACELP and the AF-CELP on TMS320C5X DSP

		8 kbit/s CS-ACELP	64 kbit/s AF-CELP
Complexity (MIPS)	Worst Case	34	29
	Minimum	29	20
	Average	32	22
ROM (kW)	Program	15.90	16.17
(16 bits)	Data	3.23	3.23
RAM (kW) (16 bits)		5.75	6.12

Ⅳ. Conclusion

A new AF-CELP coder was proposed by incorporating the adaptive fixed codebook (AFC) concept into CELP coder. By combining the adaptive codebook search, the AFC search can be done in a form of sequential optimum, joint optimum, and hybrid optimum search. Each search method should be carefully chosen by trading off the computational complexity and the coder performance. Listening test shows that an AF-CELP operating at 6.4 kbit/s achieves the comparable speech quality to the CS-ACELP at 8 kbit/s in real situations such as babble noise, car noise, and interfering talker environment. A variable rate coder switching from the CS-ACELP and the AF-CELP every 10 ms was also designed and it is shown that the performance of the variable rate coder is equivalent to that of the CS-ACELP.

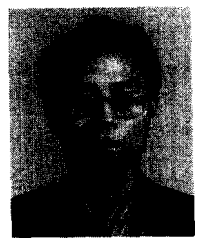
V. Acknowledgement

The author would like to thank Nam Kyu Ha for the real-time implementation of the proposed speech coder and wishes to thank the voluntary participants from human & computer interaction Lab. and digital communication Lab., Samsung Advanced Insititute of Technology for listening tests. The author also would like to thank the anonymous reviewers whose valuable comments and questions led to the improvement of this paper.

References

1. A. Gersho, "Advances in speech and audio compression," *Proc. of the IEEE*, Vol. 82, No. 6, pp. 900-918, June 1994.
2. R. Salami, C. Laflamme, J-P. Adoul, and D. Massaloux, "A toll quality 8 kb/s speech codec for the personal communications system (PCS)," *IEEE Trans. Veh. Technol.*, Vol. 43, No. 3, pp. 808-816, Aug. 1994.
3. H. K. Kim, Y. D. Cho, M. Y. Kim, and S. R. Kim, "A 4 kbit/s renewal code-excited linear prediction speech coder," in *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing*, Munich, Germany, Vol. 2, pp. 767-770, Apr. 1997.
4. K. Mano, T. Moriya, S. Miki, H. Ohmuro, K. Ikeda, and J. Ikedo, "Design of a pitch synchronous innovation CELP coders for mobile communications," *IEEE J. Select. Areas Commun.*, Vol. 13, No. 1, pp. 31-41, Jan. 1995.
5. M. E. Perkins, K. Evans, D. Pascal, and L. A. Thorpe, "Characterizing the subjective performance of the ITU-T 8 kb/s speech coding algorithm-ITU-T G.729," *IEEE Commun. Mag.*, Vol. 35, No. 9, pp. 74-81, Sept. 1997.
6. ITU-T Revised Recommendation P.800, "Methods of subjective determination of transmission quality," 1996.
7. ITU-T User's Group on Software Tools, "ITU-T software tool library manual," Geneva, May 1996.
8. R. Salami, C. Laflamme, B. Bessette, and J-P. Adoul, "ITU-T G.729 Annex A: Reduced complexity 8 kb/s CS-ACELP codec for digital simultaneous voice and data," *IEEE Commun. Mag.*, Vol. 35, No. 9, pp. 56-63, Sept. 1997.

▲Hong Kook Kim



Dr. Hong Kook Kim was born in Seoul, Korea on October 24, 1965 and received a B.S. degree in control and instrumentation engineering from Seoul National University, Korea in 1988. He received an M.S. degree and a Ph.D. degree in electrical engineering from Korea Advanced Institute of Science and Technology, Korea in 1990 and 1994, respectively.

He is working with Samsung Advanced Institute of Technology (SAIT) as a member of technical staff. During the Ph.D. candidate, he partly worked with Samsung Electronic Co. Ltd. from 1991 to 1993 and SAIT from 1993 to 1994, developing the voice dialing system for digital keyphone and the DSP implementation of speech synthesizer. From the last of 1994 in SAIT, he has developed a speech coder especially operating at 4 kbit/s. In 1996, he was awarded the best patent prize for proposing a new speech coder from SAIT. Currently, he joined the security team of SAIT in order to apply security algorithms to speech coders and computer networks. His interests include speech analysis, speech recognition and coding, and cryptosystems.