

CELP 음성부호화기 전송률 감소를 위한 음성신호의 V/UV 결정 알고리즘

The V/UV Decision Algorithm for a Reduction of the Transmission Bit Rate in the CELP Vocoder

민소연*, 김현철**

So-Yeon Min* and Hyun-Chul Kim**

요 약

기존의 CELP 계열 보코더에서, 무성음에 대한 별도의 처리 없이 유성음과 동일하게 처리함으로써 합성음에서 음질의 저하 및 계산량과 전송률 측면에서 손실을 가져왔다. 본 논문에서는 에러율과 전처리 계산량을 최소로 할 수 있는 V/UV 분류기를 사용하여 CELP 보코더에서 전송률을 감소시키는 방법을 제안한다. 새로운 V/UV 분류기는 CELP 보코더에서 사용되는 LSP 파라미터의 주파수영역 분포도와 간격정보를 이용하여 V/UV를 결정하게 된다. 제안한 방법을 G.723.1 5.3kbps ACELP에 적용하여 성능 평가를 실시하였다. 실험결과, 음질의 저하 없이 약 6%의 전송률을 감소할 수 있었다.

Abstract

The conventional CELP(code excited linear prediction) type vocoder has no V/UV(voiced/unvoiced) classifier. So, the unvoiced speech is processed like the voiced speech. In this paper, to reduce the bit rate, we propose a new V/UV decision algorithm minimized error rate and preprocessing computation. This V/UV classifier use the LSP(line spectrum pair) parameter which is acquired spectrum analysis process in CELP vocoders. Applying this method to the 5.3kbps ACELP(algebraic code excited linear prediction) in the G.723.1, we can get the transmission bits rate reduction of 6% approximately without degradation of speech quality.

Key words : CELP vocoder, V/UV(voiced/unvoiced) classifier, LSP(line spectrum parameter)

I. 서 론

과거의 음성 정보를 전송하는 방법은 한정된 일부 사용자와 특정 분야에 적용되었기 때문에 전송률은 크게 고려되지 않았으며 수신단에 고음질의 음성을 보낼 수 있으면 되었다. 그러나 멀티미디어와 이동

통신의 발달로 인해 특정 집단이나 개인에게 제공되었던 서비스들이 이제는 다수의 대중에게도 제공하게 되었고 그 숫자 역시 기하 급수적으로 늘어나게 되었다. 그로 인해 지금까지 적용되어왔던 전송률로는 사용자 집단의 숫자를 충족시킬 수 없게 되었다. 또한 전송률을 저하시켜 동일 채널 상에 사용자 수를

* 서일대학 정보통신과 (Dept. of Information & Communication, Seoil College)

** 숭실대학교 컴퓨터학과 (Dept. of Computer Graduate School, Soong-sil University)

· 제1저자 (First Author) : 민소연

· 접수일자 : 2006년 12월 18일

증가시킬 경우에 발생하는 음질 열화 역시 문제가 되었다. 이와 같은 배경 하에서 음성 부호화기, 즉 보코더의 개발이 이루어지게 되었다[1,2].

CELP형 보코더는 낮은 전송률에 비해서 우수한 음질을 제공하고 있다. 하지만, 무성음에서도 유성음과 같은 분석을 통해서 데이터를 전송하므로 보다 많은 계산량이 필요하게 되고 전송률 측면에서도 손실을 가져오게 된다. 또 V/UV 분류기를 사용하는 경우 이것의 성능에 따라 음질 저하의 차이가 심하다.

그러므로 본 논문에서는 기존의 CELP 보코더에서 사용되는 LSP 파라미터의 특성을 이용하여 유성음과 무성음을 분리하여 전송률을 낮추는 방법을 제안하고자 한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 유,무성음의 LSP 파라미터의 분포특성을 설명하고 3장에서는 제안한 유,무성음 성분분리 알고리즘에 대해 설명한다. 4장에서는 시뮬레이션을 통해 제안한 알고리즘의 분석 결과를 살펴보고 5장에서는 결론을 맺는다.

II. 유·무성음의 LSP 파라미터 분포 특성

LSP 파라미터는 일정한 스펙트럼 민감도와 저전송률 부호화 특성을 지니며 낮은 스펙트럼 왜곡, 좋은 선형보간 특성을 가진다. LSP 파라미터를 얻기 위해서는 먼저 선형예측방법에 의해 LPC(linear prediction coefficient)를 구하고 이를 변환하여야 한다. 변환방법은 복소근(complex root) 방법, 실근(real root) 방법, ratio filter 방법 등이 있으나 실근 방법이 주로 사용된다[3,4,7].

그림 1은 음성신호의 무성음과 유성음에 대한 스펙트럼과 LPC 분석 그리고 LSP 파라미터를 보여주고 있다. 유성음의 스펙트럼은 (b)와 같이 1kHz 이내에서 제 1 포먼트가 나타나고 3개 이상의 포먼트가 존재한다. 반면에 무성음의 스펙트럼은 (a)와 같이 저주파수 영역의 에너지가 낮고 고주파수 영역의 에너지가 높다. 이와 같은 스펙트럼모양으로 LSP의 모양도 유성음과 무성음에서 뚜렷한 차이를 가진다. 먼저 유성음에서는 포먼트로 인해 저주파수 영역에서 보다 많은 LSP가 분포하고 그 간격도 고주파수의 LSP

간격 보다 좁게 나타난다. 반면에 무성음에서는 고주파수 영역에서 LSP가 많이 분포하고 그 간격도 저주파수영역의 LSP 간격보다 좁게 나타난다[4, 6~8].

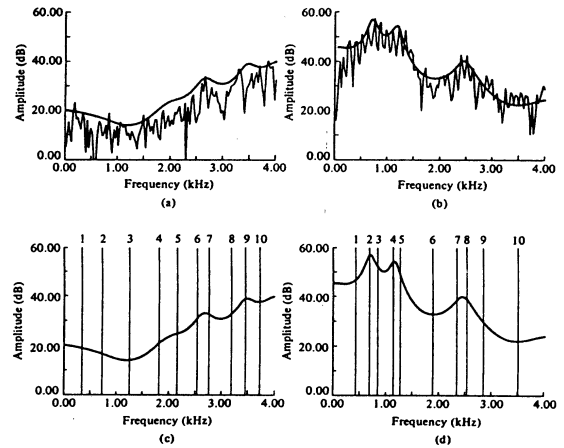


그림 1. LSP 파라미터의 예
 (a) 자음 /s/ (b) 모음 /a/
 (c),(d) /s/와 /a/에 대한 LPC 분석과 LSP
 Fig.1 Example of the LSP parameter
 (a) Consonant /s/ (b) Vowel /a/
 (c),(d) The analysis of LPC and LSP for /s/,/a/

III. 음성신호의 성분분리를 통한 전송률 감소 알고리즘

3-1 LSP 파라미터를 이용한 신호의 성분분리

2장의 유,무성음의 LSP 파라미터 분포 특성에서 기술한 바와 같이 유,무성음에 따른 LSP 파라미터의 특성으로 인해 음성신호의 특성이 차이가 난다. 이에 본 논문에서는 먼저 유,무성음의 LSP 분포 차이를 이용한다. 음성신호의 성분분리를 위해 샘플링 주파수를 F_s 라고 하고 $F_s/4$ 이하의 주파수영역에 존재하는 LSP 개수를 NL로 나타내었다. 또한, $F_s/4$ 이상의 주파수영역에 존재하는 LSP 개수를 NH로 나타내어, NL이 NH보다 큰 경우는 음성신호의 스펙트럼이 저주파 쪽에서 봉우리(pole)가 많이 나타나는 모양이어서 유성음의 스펙트럼 특징을 나타낸다고 판단한다. 왜냐하면 유성음의 제 1 포먼트와 제 2 포먼트가 주로 저주파수 영역에 존재하기 때문이다.

이와 반대로 NH가 NL보다 큰 경우는 무성음을 나타낸다고 결정한다. 그 이유는 무성음의 스펙트럼은

주된 포먼트가 고주파영역에 나타나기 때문이다. 하지만 /i/, /l/, /e/, /æ/ 와 같은 유성음은 제 2 포먼트, 제 3 포먼트 또는 제 4 포먼트가 고주파수 영역에 존재하여 유성음이면서도 NH가 NL보다 크게 나타난다. 이와 같은 경우에는 제 1 포먼트의 존재 여부로서 무성음인지 유성음인지를 결정하게 된다. 즉, LSP 파라미터들의 간격을 조사하여 $F_s/4$ 이하의 영역에서 좁은 간격을 나타내는 LSP 들이 존재하면 유성음으로 간주한다. 앞에서 설명되어진 절차에 의하여 LSP 파라미터의 분포특성을 사용하여 유,무성음 성분분리를 나타낸 결과를 그림 2와 그림 3에 나타내었다.

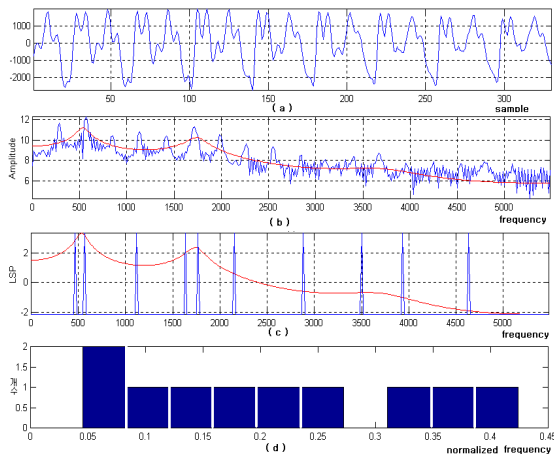


그림 2. 유성음에 대한 LSP 분포
(a)유성음 파형 (b)유성음 파형의 주파수 스펙트럼
(c)LPC 분석과 LSP 파라미터 (d)LSP 분포특성

Fig. 2 LSP distribution of voiced speech signal

- (a) Waveform of voiced speech signal
- (b) Spectrum of voiced speech signal
- (c) LPC analysis and LSPs
- (d) Distribution of LSPs

그림 2는 유성음에 대한 LSP 분포를 나타낸다. (d)의 histogram에서 10개의 막대 중 앞에서 5개의 막대가 NL을 구성하는 것이고 나머지 5개의 막대가 NH를 구성하는 것이다. NL이 NH보다 큰 것을 알 수 있어 histogram만으로도 유성음임을 알 수 있다. 반면에 그림 3의 (d)에서는 NH가 NL보다 크게 나타남을 알 수 있다. 즉, 유성음에서는 LSP가 저주파수 영역에 많이 분포하고 무성음에서는 고주파수 영역에 많이 분포함을 알 수 있다.

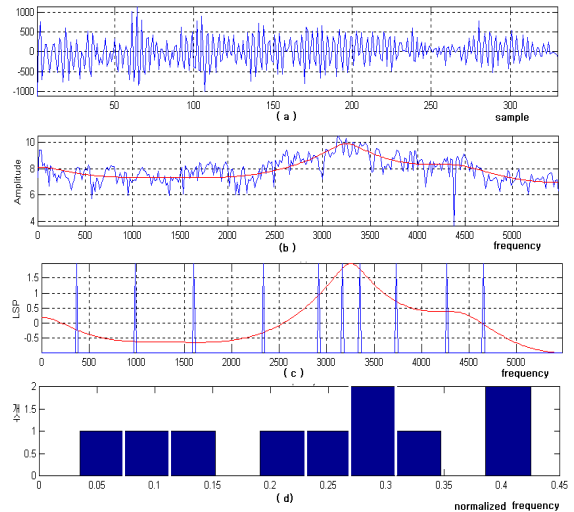


그림 3. 무성음에 대한 LSP 분포
(a) 무성음 파형 (b) 음성파형의 스펙트럼
(c) LPC 분석과 LSP (d) LSP 분포

Fig. 3 LSP distribution of unvoiced speech signal

- (a) Waveform of unvoiced speech signal
- (b) Spectrum of unvoiced speech signal
- (c) LPC analysis and LSPs
- (d) Distribution of LSPs

3-2 성분분리를 통한 전송률 감소 알고리즘

성분분리를 통한 전송률 감소 알고리즘에 대한 전체 블록도를 그림 4에서 나타내었다.

입력신호에서 묵음부분과 음성부분을 분리과정은 기존의 G.723.1 ACELP에서 사용하는 VAD(voice activity detection)를 사용하였다. 이것으로 검출된 음성부분에 대해서만 유,무성음 결정을 수행하였다. VAD가 1이면 음성부분인데 이런 구간에 대해 먼저 NL과 NH를 비교하여 NL이 크면 유성음으로 간주한다. 그렇지 않으면 LSP 파라미터 중 간격이 최소가 되는 주파수 영역을 찾는다.

그림 4에서 $\min(intv)$ 는 LSP 파라미터들의 간격 중 최소가 되는 주파수를 나타낸다. 여기서 $intv$ 는 LSP 파라미터들의 간격정보를 나타낸다.

$$P = [p_1, p_2, p_3, \dots, p_{10}] \tag{1}$$

$$intv(i) = |p_{i+1} - p_i| \quad , \quad 1 \leq i \leq 9$$

P는 LSP 파라미터를 나타내므로, $intv$ 는 간격정보

를 의미하게 된다.

$\min(intv)$ 가 $F_s/4$ 보다 작으면 제 1 포먼트가 저주파수 영역에서 위치한 것으로 결정하여 $NL < NH$ 이지만 유성음 구간으로 판단한다. $\min(intv < F_s/4)$ 의 조건은 $F_s/4$ 의 주파수 영역에서 나타나는 LSP의 $intv$ 중 가장 작은 것을 의미하고 $\max(intv < F_s/4)$ 는 가장 큰 것을 의미한다.

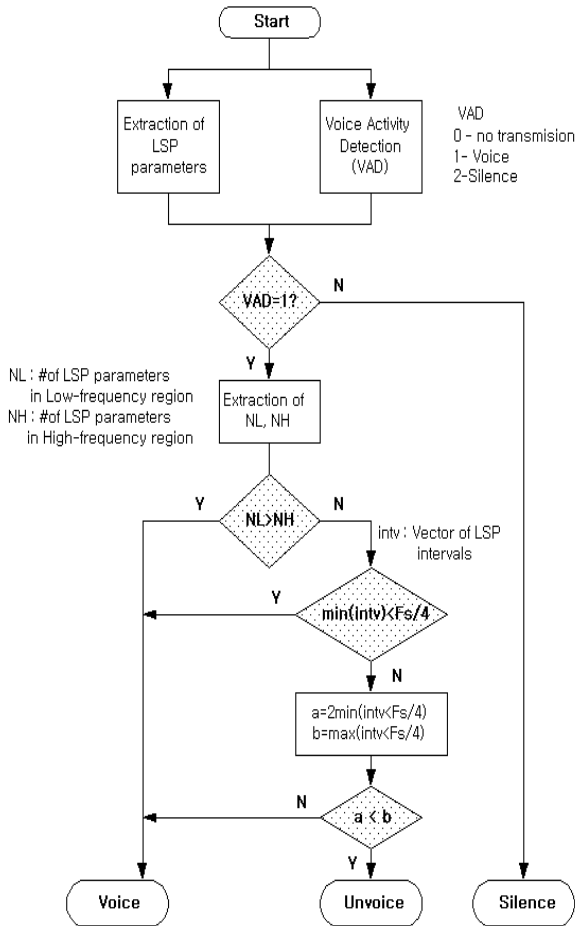


그림 4. 음성신호의 U/V 결정 알고리즘
Fig. 4 U/V decision algorithm of speech signal

$\min(intv < F_s/4)$ 의 두 배를 a 라 하고 $\max(intv < F_s/4)$ 를 b 라 한다면 $a < b$ 이면 $F_s/4$ 이하의 주파수 영역에서 좁은 간격의 LSP 파라미터들이 존재함을 반영하는 것으로 판단하여 유성음으로 결정하고 $a > b$ 이면 무성음으로 결정한다.

위에서 설명되어진 LSP 파라미터 분포에 의한 유, 무성음 결정에 대해 각기 다른 부호화 방법을 적용하여 전송률을 줄이게 된다. 즉, 유성음은 기존의 방식

을 사용하여 부호화 하고 무성음에 대해서는 LSP 파라미터와 gain만 전송한다.

IV. V/UV 결정 알고리즘의 성능 시뮬레이션

LSP 파라미터를 이용한 성분분리 알고리즘을 C언어로 구현하여 5.3kbps ACELP (ITU-T 표준안 G.723.1) 보코더에 적용하였다. 실험에 사용한 보코더의 전체 블록도를 그림 5에서 나타내었다.

프레임 크기는 30msec이고 부프레임은 5msec이다. LSP는 한 프레임을 4개의 부 프레임으로 나누고 각 부프레임별 LSP를 6bit VQ로 양자화하여 24bit만 전송하고 신호에 대한 gain을 6bit 전송한다. 그리고 이를 제어할 비트를 2bit 전송한다. 이렇게 하여 무성음 프레임에 대해서는 32비트를 전송한다. 기존의 G.723.1에서는 무성음에 대해서도 158bit를 전송하는데 비해 126bit를 줄이게 된다.

제안한 알고리즘의 성능 평가를 위해 전송률 비교, 부호화기를 통과한 합성음의 음질 비교, 합성음에 대한 과형 비교를 분석하여 평가하였다.

음성신호의 음질 평가 방법에는 객관적 음질 평가 방법인 SNR(signal to noise ratio) 비교와 주관적 음질 평가방법인 MOS(mean opinion score) test 방법이 있다. 두 가지 방법 중에서 MOS test 방법을 사용하여 음질 열화의 발생여부를 측정하였다. 표 1에서는 G.723.1의 VAD만 사용한 보코더의 전송률과 LSP 파라미터를 이용하여 성분분리를 수행한 보코더의 전송률을 비교하였다. 표 2에서는 기존의 방법과 제안한 방법 간의 음질을 비교하였다. 전송률은 제안한 방법이 기존의 5.3kbps ACELP보다 약 6% 감소하였고 음질열화는 발생하지 않았다. 그림 6에서는 기존의 ACELP 부호화기를 통과한 합성음 과형과 제안한 유,무성음 결정 알고리즘을 적용한 음성부호화기를 통과한 합성음에 대한 비교 결과를 나타내었다. 실험 결과를 통해 기존의 방법과 같은 합성음 과형 결과를 얻을 수 있었다.

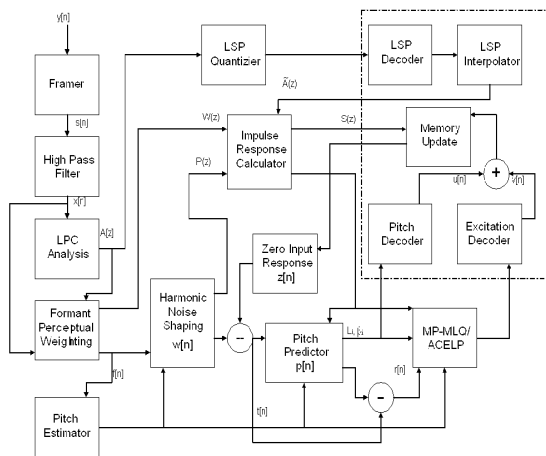


그림 5. G.723.1 부호화기의 블록도
Fig. 5 Block diagram of G.723.1 vocoder

표 1. 전송률 감소에 대한 실험결과

Table 1. Experimental result of transmission bit rate reduction

	G.723.1 (5.3kbps- ACELP)	Proposed Method	Experimental Result
발성 1	5.251	4.991	0.260
발성 2	4.656	4.447	0.209
발성 3	5.044	4.724	0.320
발성 4	4.999	4.670	0.329

표 2. 기존 방법과 제안한 방법의 MOS 비교

Table 2. Experimental result of MOS test

	G.723.1 (5.3kbps)	Proposed Method
발성 1	3.84	3.8
발성 2	3.76	3.7
발성 3	3.8	3.78
발성 4	3.7	3.7

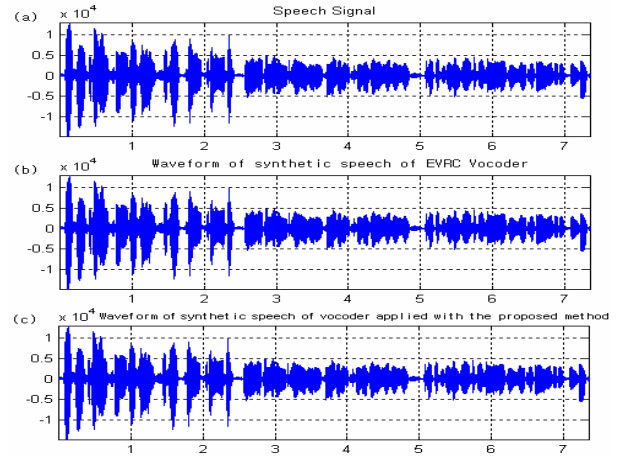


그림 6. 음성부호화기를 통과한 합성음의 비교결과
(a) 원 음성신호
(b) ACELP 음성부호화기의 합성음
(c) 제안한 알고리즘을 적용한 음성부호화기의 합성음
Fig. 6 Comparison of synthetic speech of encoding device.

- (a) Original voice signal
- (b) Waveform of synthetic speech of ACELP voice encoding device
- (c) Waveform of synthetic speech of encoding device applied with the suggested method

V. 결 론

음성부호화 시 처리될 음성 신호 구간이 유성음, 무성음 또는 묵음부분인지의 정보를 통해 중요도를 달리하거나 각 성분에 보다 적합한 파라미터를 사용하는 방법을 통해 비트를 줄이거나 음질을 향상시킬 수 있다. 따라서 음성신호처리에서 성분분리의 전처리과정은 매우 중요하다.

본 논문에서는 CELP 보코더에서 LP(linear prediction) 분석시 얻어지는 LSP 파라미터를 사용하여 유,무성음을 결정하여 전송률을 줄일 수 있는 방법을 제안하였다. 저주파수 영역과 고주파수 영역의 LSP 분포와 포먼트의 존재여부를 결정하는 LSP 간격정보를 이용하여 음성신호의 성분을 분리하였다. 이렇게 분리된 무성음 구간에 대해 LSP 파라미터와 gain정보만 전송함으로써 전송률을 낮출 수 있었다.

CELP형 보코더인 G.723.1의 5.3kbps(ACELP)에 유,무성음 결정 알고리즘을 적용한 결과, 음질의 열화는 발생하지 않았고 약 6%의 전송률을 감소시킬 수 있었다.

감사의 글

본 논문은 2006년도 서일대학 학술연구비에 의해 연구되었습니다.

참 고 문 헌

- [1] Mabo Robert Ito, Robert W. Donaldson, "Zero-Crossing Measurements for Analysis and Recognition of Speech Sounds", *IEEE Trans. on A.A.E.* Vol. AU-19, No. 3 pp. 235-242, Sep., 1971.
- [2] B.S Atal, L.R.Rabiner, "A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Application to Speech Recognition", *IEEE Trans. on ASSP*, Vol. ASSP-24, No. 3. June, 1976.
- [3] A. M. Kondoz, *Digital Speech, John Wiley & Sons Ltd*, 1994.
- [4] H. Kobatake, "Optimization of Voiced/Unvoiced Decisions in Noise Environments", *IEEE Trans. on ASSP*, Vol. ASSP-35, No. 1, pp. 9-18, Jan., 1987.
- [5] S.G. Knorr, "Reliable Voiced/Unvoiced Decision", *IEEE Trans. on ASSP*. Vol. ASSP-27, No. 3, June, 1979
- [6] L.R. Rabiner and R.W. Schafer, "Digital processing of Speech Signals Englewood Cliffs", New Jersey : Prentice-Hall, 1978.
- [7] 민소연, 배명진, "LSP 파라미터 분포 특성을 이용한 주파수 대역 조절법에 대한 연구", *한국음향학회지*, 제21권, 제3호, 2002년 4월.
- [8] 민소연, 장경아, 배명진, "선형 스펙트럼쌍을 이용한 성문특성이 제거된 성도특성 추출법에 관한 연구", *한국음향학회지*, 제21권 7호, 2002년 10월

민 소 연 (閔素娟)



1993년 2월 : 숭실대학교 전자공학과(공학사)
 1995년 2월 : 숭실대학교 전자공학과(공학석사)
 2003년 2월 : 숭실대학교 전자공학과(공학박사)

2005년 3월 ~ 현재 : 서일대학 정보통신과 교수
 관심분야 : 통신 및 신호처리, 통신 알고리즘, 네트워크 보안, 멀티미디어 신호처리 등

김 현 철 (金賢鐵)



2003년 2월 : 인제대학교 정보컴퓨터학부(공학사)
 2005년 2월 : 경원대학교 전자계산학과(공학석사)
 2006년 3월 ~ : 숭실대학교 컴퓨터학과 박사과정

관심분야 : PKI, Home-Network 보안, OFDM 등