

# 개선된 혼성영역 교차상관법에 의한 G.723.1의 피치검색시간 단축에 관한 연구

|          |
|----------|
| 논 문      |
| 59-12-33 |

## A Study on the Pitch Search Time Reduction of G.723.1 Vocoder by Improved Hybrid Domain Cross-correlation

조 왕 래\* · 최 성 영\*\* · 배 명 진†  
(Wang-Rae Jo · Seong-Young Choi · Myung-Jin Bae)

**Abstract** - In this paper we proposed a new algorithm that can reduce the open-loop pitch estimation time of G.723.1. The time domain cross-correlation method is simple but has long processing time by recursive multiplication. For reduction of processing time, we use the method that compute the cross-correlation by multiplying the Fourier value of speech by it's complex conjugate. Also, we can reduce the processing time by omitting the bit-reversing of FFT and IFFT for time-frequency domain transform. As a result, the processing time of improved hybrid domain cross-correlation algorithm is reduced by 67.37% of conventional time domain cross-correlation.

**Key Words** : CELP, G.723.1, Open-loop pitch estimation, Cross-correlation, Bit-reversing

### 1. 서 론

정보통신의 비약적 발전에 따라 무선 이동통신을 통한 서비스의 범위도 과거의 음성 통신에서 영상, 데이터, 멀티미디어 등으로 점차 확대되고 있다. 하지만 휴대전화의 기본 기능은 여전히 음성통신에 있다고 볼 수 있으며, 이를 위해 디지털 이동통신 분야와 인터넷 기반의 멀티미디어 전송에 적용하기 위한 음성신호의 디지털 변환과 전송 데이터량을 줄이기 위한 음성 부호화 기술에 대한 연구가 진행되고 있다. 또한, ITU, ETSI 등의 표준화 단체들에서는 이러한 음성 부호화 기술에 대한 표준화 작업이 활발히 진행되고 있다.

표준 음성 부호화기는 1972년에 ITU-T G.711로 채택된 64kbps PCM 방식을 시작으로 32kbps ADPCM, 16kbps LD-CELP 방식으로 표준화 되었다. 현재 ITU-T에서는 PCS, IMT-2000 등에서 사용할 수 있는 8kbps 음성 부호화기에 대한 표준화 작업으로 1996년에 CS-CELP를 G.729로 인터넷폰 및 화상통신용 음성 부호화기로 ACELP/MP-MLQ의 5.3/6.3kbps 이중 속도 부호화 방식을 G.723.1로 선정하였다.

G.723.1에서는 음성 생성 모델을 기반으로 성도모델은 LPC 계수로, 성대 여기 모델은 피치지연값으로 부호화하고 이 두 필터에 의하여 모델링되지 않은 값들에 대해서는 코드북을 사용하여 보정한다.

본 논문에서는 G.723.1의 피치값을 추정하는 시간을 단축하기 위하여 표준에 사용되고 있는 시간영역 교차상관법 대

신에 개선된 시간-주파수 혼성영역 교차상관법을 제안한다. 혼성영역 교차상관법의 계산과정에 사용되는 FFT와 IFFT의 비트-재정렬 과정을 생략함으로써 처리시간을 크게 단축할 수 있게 되었다.

### 2. G.723.1 음성부호화기

G.723.1 음성부호화기는 선형예측 합성에 의한 분석 부호화를 기본으로 사용하며 인지 가중된 오차 신호를 최소화하도록 부호화한다. 입력된 음성에 대하여 240샘플(8kHz 샘플링율에서 30ms)을 한 프레임으로 처리하며 DC 성분을 제거하기 위해 하이패스필터를 통과시키고 60샘플 단위의 4개의 서브 프레임으로 나눈다. 각각의 서브 프레임에 대해 10차의 LPC를 계산하며 양자화되지 않은 LPC 계수들을 이용하여 단구간 인지 가중 필터를 구성하고 프레임에 적용하여 인지 가중화된 음성신호를 얻는다.

두 서브프레임(120샘플)에 대하여 인지 가중화된 음성신호를 사용하여 개회로 피치 주기를 추정하며 피치 주기는 18 ~ 142 샘플 구간에 대해서만 수행한다. 이전에 계산된 피치 추정값을 사용하여 noise shaping 필터를 구성하고 LPC 합성필터, 포먼트 인지 가중 필터와 결합하여 임펄스 응답을 계산한다.

임펄스 응답과 개회로 피치 추정값은 폐회로 피치 예측값이 계산된다. 폐회로 피치예측은 개회로 피치추정값 주변의 작은 편차에 대하여 추정하고 피치 주기값과 차분값이 복호화기로 전송된다.

마지막으로 여기신호의 비주기적인 성분을 추정하는데 높은 전송율에서는 MP-MLQ(Multipulse Maximum Likelihood Quantization) 여기가 사용되고 낮은 전송율에서는 ACELP(Algebraic-Code excitation)가 사용된다[1]. G.723.1 부호화기의 블록다이어그램을 그림 1에 나타내었다.

\* 정 회 원 : 숭실대 공대 정보통신공학과 박사

\*\* 정 회 원 : 한국폴리텍2대학 전자과 교수

† 교신저자, 정회원 : 숭실대 정보통신공학과 교수

E-mail : mjbae@ssu.ac.kr

접수일자 : 2009년 10월 27일

최종완료 : 2010년 7월 14일

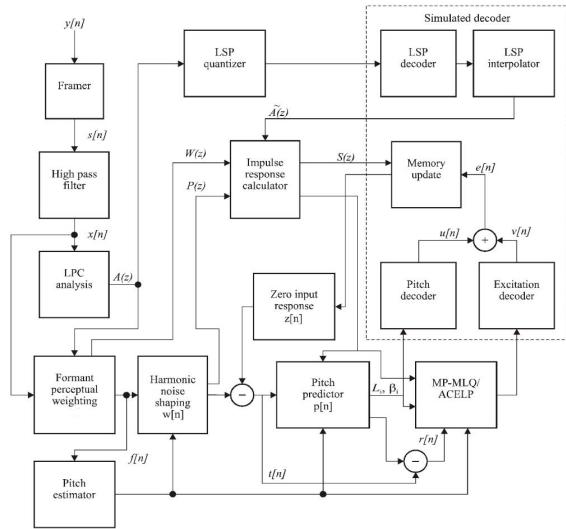


그림 1 G.723.1 부호화기의 블록도  
Fig. 1 Block diagram of G.723.1 encoder

### 3. G.723.1의 피치추정법

G.723.1 음성 부호화기는 프레임당 두 번의 피치 추정을 수행한다. 한번은 앞의 두 서브 프레임에 대하여 수행하고 나머지 한번은 뒤의 두 서브 프레임에 대해서 수행한다. 개회로 피치 주기 추정값  $L_{OL}$ 은 인지 가중된 음성신호  $f(n)$ 을 사용하여 계산한다.

최대 교차상관관계값  $C_{OL}(j)$ 은 피치 주기를 결정하기 위해 다음과 같이 계산된다.

$$C_{OL}(j) = \frac{\left( \sum_{n=0}^{119} f(n) \cdot f(n-j) \right)^2}{\sum_{n=0}^{119} f(n-j) \cdot f(n-j)}, \quad 18 \leq j \leq 142 \quad (1)$$

교차상관관계값  $C_{OL}(j)$ 을 최대로 하는 인덱스  $j$ 가 해당 두 서브 프레임에 대한 개회로 피치 추정값으로 선택된다. 피치 추정시에 피치 주기의 배수값들이 찾아지는 것을 예방하기 위해 작은 피치주기 값들에 우선권이 부여된다. 피치 주기를 찾는 구간은 인덱스  $j$ 가 18인 구간부터 찾기 시작하며 모든 최대값을 찾아 이전의 최대값  $C_{OL}(j')$ 과 비교한다. 만일 인덱스  $j$ 와  $j'$ 의 차이가 18보다 작고  $C_{OL}(j) > C_{OL}(j')$  이면 새로운 최대값  $j$ 가 선택되며 인덱스간의 차이가 18보다 크거나 같다면  $C_{OL}(j)$ 이  $C_{OL}(j')$  보다 1.25dB 이상 클 경우에만 새로운 최대값이 선택된다[4].

### 4. 개선된 혼성영역 교차상관법

음성신호는 시간영역에서 여기성분과 여파기성분의 컨벌루션으로 식 (2)와 같이 나타낼 수 있다.

$$s(n) = e(n) * h(n) \quad (2)$$

이를 주파수 영역으로 변환하면 식 (3)과 같이 나타낼 수 있다.

$$S(K) = E(K) \cdot H(K) \quad (3)$$

이는 식 (2)와 같은 두 신호의 컨벌루션을 계산할 때 각 신호의 FFT를 구하여 곱하고 다시 IFFT하면 쉽게 구할 수 있다는 것을 의미한다[2].

음성신호의 교차상관값은 식 (4)와 같이 정의되며 이를 주파수 영역으로 변환하면 식(5)와 같다.

$$R(k) = \sum_{m=-\infty}^{\infty} x(m)y(m+k) \quad (4)$$

$$R(K) = X(K) \cdot Y^*(K) \quad (5)$$

여기서  $Y^*(K)$ 는  $Y(K)$ 의 켈레복소수이다. 즉, 각 신호의 FFT를 구하여 한쪽에 대해서는 켈레복소수를 구하여 곱한 다음 IFFT를 하면 교차상관값을 구할 수 있다[2].

자기상관함수(Autocorrelation)는 교차상관함수(Cross-correlation)의 특수한 경우이며 G.723.1의 피치추정에 사용되는 방법은 자기상관법에 해당한다. 자기상관함수는 시간 지연에 따른 상관성을 나타내는 함수로서 시간영역상에서는 주로 신호의 주기성을 찾고자 할 때 사용된다. 그림 2에 음성신호의 자기상관결과를 그래프로 나타내었고 그림 3에는 혼성영역 교차상관법에 의해 피치를 구하는 과정을 블록도로 나타내었다.

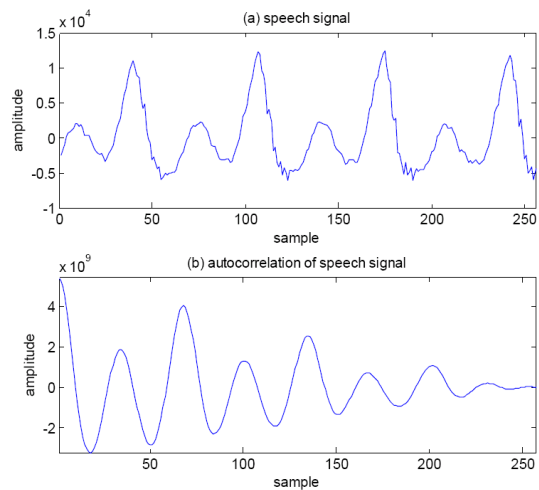


그림 2 음성신호의 자기상관값  
Fig. 2 Autocorrelation value of speech signal

그림 3에 나타낸 바와 같이 입력된 음성신호에 대하여 윈도우를 적용하고 이를 FFT한다. FFT의 결과와 그 결과의 켈레복소수를 만들어 곱한 다음 다시 IFFT를 하게되면 원신호의 자기상관값을 얻을 수 있다. 자기상관함수의 특성에 의해 피크를 구하고 피크까지의 거리를 구하면 음성신호의 피치 주기를 구할 수 있다.

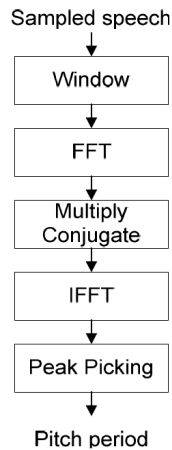


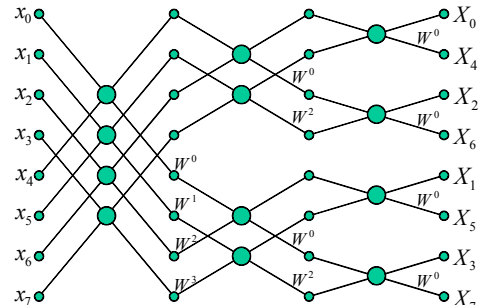
그림 3 혼성영역 교차상관법의 처리과정  
Fig. 3 Process of cross-correlation in hybrid domain

FFT는 DFT(Discrete Fourier Transform)를 계산하는데 있어 결과는 같으면서도 연산수를 줄여 계산속도를 높이는 방법이다. 계산량을 살펴보면 N개의 샘플을 DFT하는데 각 n에 대하여 N번의 복소수 곱셈이 필요하게 되어 결과적으로  $N^2$ 에 비례하는 계산량이 필요하게 된다. 그러나 N개의 샘플을 FFT하는 경우에는 같은 결과를 내면서 계산량은  $N \times \log_2 N$ 에 비례하도록 줄일 수 있다[2].

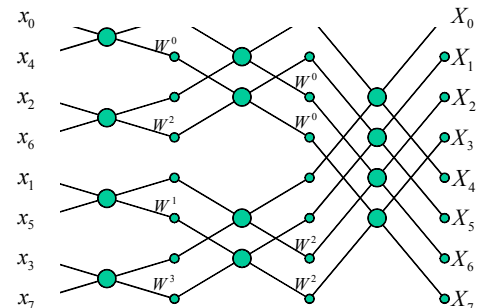
FFT는 DIT(decimation in time)와 DIF(decimation in frequency) 각각에 대해 정상 순서의 입력을 사용한 경우와 비트-재정렬된 입력을 사용하는 방법이 있다. FFT 알고리즘에 가장 많이 사용되는 Cooley-Tukey 알고리즘은 DIF 방법을 사용하며, IFFT의 경우에는 FFT와 같은 방법을 사용하면서 단지 계수들의 켤레 복소수(complex conjugate)를 사용하고 루틴의 끝에서  $1/N$  스케일링(scaling)을 수행하는 것만이 다르다[2]. 그러나 FFT는 계산하고자하는 데이터 샘플 수가  $N=2^v$  ( $v$ 는 정수)가 되어야 한다는 것과 그림 4(a)에 나타난 바와 같이 입력배열과 출력배열의 순서가 서로 일치하지 않는다는 단점이 있다. 따라서 FFT 수행 전이나 수행 후에 배열의 순서를 재정렬해 주어야만 한다. 이를 비트-재정렬(bit-reversing)이라 하며 계산량에 있어 큰 오버헤드로 작용하게 된다. 이러한 오버헤드는 적은 샘플수를 갖는 데이터에 대한 FFT 연산이 DFT에 비해 큰 이점이 없도록 하며 캡스트럼 분석과 같이 시간-주파수 영역 변환이 잦은 연산의 처리속도에 큰 영향을 미치게 된다[2].

본 논문에서는 G.723.1 음성 부호화기의 피치 추정 시간을 단축하기 위하여 기존의 시간영역 교차상관관계법을 사용하는 대신 개선된 시간-주파수 혼성영역 교차상관법을 이용하여 피치 주기를 추정하는 방법을 제안하였다. 시간영역 교차상관법의 피치 추정시간을 개선하기 위하여 FFT와 IFFT를 이용하는 피치 추정법을 제안하였으며 영역 변환에 사용되는 FFT와 IFFT의 비트-재정렬 과정을 생략함으로써 처리시간을 단축하는 방법을 제안하였다. 기존에는 FFT와 IFFT에 동일한 알고리즘을 적용함으로써 필연적으로 비트-재정렬 과정을 수행하여야 하였으며 이러한 오버헤드는 처리시간에 큰 영향을 주게 된다[2]. 그러나 FFT에는 그림 4(a)의 정상순서 입력의 DIF 방법을 사용하고 그 결과의 켤

레복소수를 구하여 곱한 후에 그림 4(b)의 비트-재정렬된 입력의 DIF 방법을 사용하여 IFFT하면 정상순서의 교차상관값을 얻을 수 있게 되어 FFT와 IFFT 과정에서 비트-재정렬 과정을 생략 할 수 있게 된다.



(a) 정상순서입력의 DIF



(b) 비트-재정렬 입력의 DIF

그림 4 8포인트 FFT의 처리흐름도  
Fig. 4 Flow graph for 8-point FFT

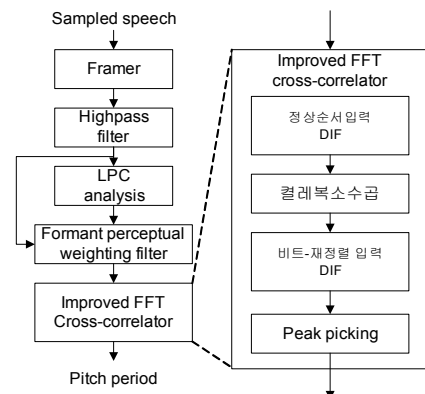


그림 5 제안한 G.723.1 피치 추정 알고리즘  
Fig. 5 Proposed pitch estimation algorithm of G.723.1

그림 5에는 제안한 알고리즘을 G.723.1 음성 부호화기에 적용하였을 때의 처리과정을 나타내었다. 입력된 음성을 프레임 단위로 나누어 DC 성분 제거를 위한 하이패스필터에 통과시킨다. 이 신호는 인지 가중 필터를 거쳐 피치 추정을 하게 된다. 피치 추정에는 혼성영역 교차상관법을 사용하며 계산과정에 사용되는 FFT와 IFFT에서 비트-재정렬 과정을 생략함으로써 처리시간을 단축할 수 있게 된다.

5. 실험 및 결과

논문에서 제안한 방법의 성능을 측정하기 위해 기존의 교차상관관계 추정 알고리즘과 개선된 혼성영역 교차상관법 알고리즘을 Lenovo/T-61p(Intel Core2 Duo 2.6GHz) 노트북에서 C++로 구현하였다. 각각의 함수는 부동소수점 연산을 사용하였고 반복하여 사용되는 회전인자(twiddle factor)인 사인값과 코사인값, 비트-재정렬 테이블은 미리 계산하여 저장하여 사용하였다. FFT와 IFFT 프로그램은 참고문헌 5와 6에 소개된 프로그램을 기본으로 사용하였다[5][6].

먼저 FFT 알고리즘에서 비트-재정렬 시간이 차지하는 비율을 알아보기 위하여 음성 신호 128샘플, 256샘플, 512샘플 단위로 FFT를 수행하면서 전체 처리시간과 비트-재정렬 시간을 측정하여 표 1에 나타내었다. 표 1에 나타낸 바와 같이 256샘플 FFT의 경우 전체 처리시간의 21.98%가 비트-재정렬에 소요됨을 알 수 있다. 따라서 제안한 알고리즘은 전통적 FFT 교차상관법보다 20%정도의 처리시간 개선이 가능하다.

다음으로 G.723.1에서 사용하고 있는 시간영역 교차상관법을 개선된 혼성영역 교차상관법 대체함으로써 얻어지는 피치 추정 속도의 개선 정도를 측정하기 위하여 시간영역 교차상관법과 혼성영역 교차상관법, 개선된 혼성영역 교차상관법의 피치추정시간을 측정하여 비교하였다. 시간영역 교차상관법은 피치추정 단위인 두 서브프레임(120샘플)에 대하여 지연값 18부터 142까지 계산하였고, 혼성영역 교차상관법과 개선된 혼성영역 교차상관법은 FFT의 특성상 256샘플을 기본단위로 수행하였다. 혼성영역 교차상관법의 FFT와 IFFT에는 정상순서 입력의 DIF방법을 사용한 후 비트-재정렬을 수행하였고 개선된 혼성영역 교차상관법은 FFT를 할 때는 정상순서 입력의 DIF방법을 사용하고 IFFT에는 비트-재정렬된 입력의 DIF방법을 사용함으로써 비트-재정렬 시간을 단축하였다. 각각의 처리시간은 1,000회씩 측정하여 평균하여 표 2에 나타내었다. 표 2의 결과에 나타낸 바와 같이 기존의 시간영역 교차상관법에 의한 피치 추정 시간이 297.77 $\mu$ s인데 비하여 개선된 혼성영역 교차상관법이 200.62  $\mu$ s로 처리샘플수가 늘어남에도 불구하고 처리시간은 67.37%로 단축됨을 알 수 있다. 그림 6에는 개선된 혼성영역 교차상관법에 의한 교차상관값을 그래프로 나타내었다. FFT의 특성상 128샘플씩 서로 대칭으로 나타나는 것을 알 수 있다. 또한 그림 7에는 시간영역 교차상관법과 개선된 혼성영역 교차상관법에 의하여 /인수네 꼬마는 천재소녀를 좋아한다./라는 발성의 피치를 추정한 그래프를 나타내었다. 그림에 나타낸 바와 같이 기존의 방법이나 개선된 방법에 의해 추정된 피치 곡선은 거의 유사한 정확도로 측정되었다.

표 1 FFT 처리시간과 비트-재정렬시간의 비  
Table 1 The ratio of FFT time vs. Bit-reversing time

|       | 처리시간( $\mu$ s) |              | 비트-재정렬 시간비(B/A) |
|-------|----------------|--------------|-----------------|
|       | 전체처리시간 (A)     | 비트-재정렬 시간(B) |                 |
| 128샘플 | 48.77          | 10.23        | 20.97%          |
| 256샘플 | 96.64          | 21.25        | 21.98%          |
| 512샘플 | 194.63         | 40.26        | 20.68%          |

표 2 처리시간의 비교

Table 2 Comparison of processing time

|                        | 피치추정시간( $\mu$ s) | 처리시간 단축율 |
|------------------------|------------------|----------|
| 표준 교차상관법 (120샘플)       | 297.77           | 100%     |
| 혼성영역 교차상관법 (128샘플)     | 234.37           | 78.71%   |
| 개선된 혼성영역 교차상관법 (128샘플) | 200.62           | 67.37%   |

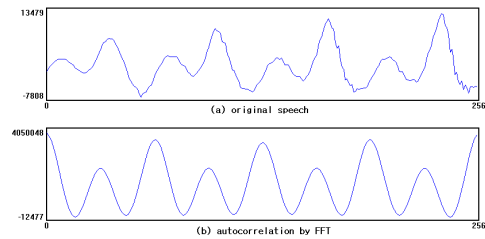


그림 6 제안한 알고리즘에 의한 교차상관값  
Fig. 6 Cross-correlation value by proposed algorithm

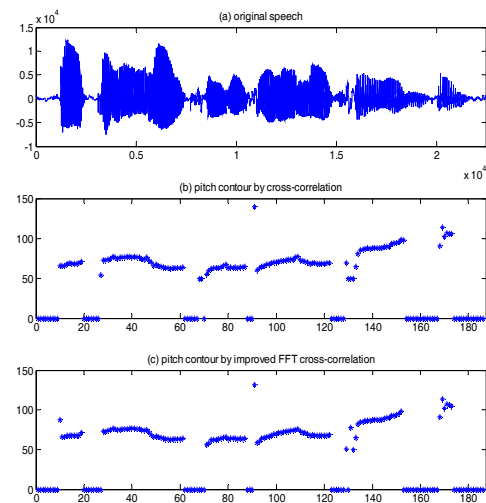


그림 7 기존의 방법과 제안한 방법에 의한 피치 곡선  
Fig. 7 Pitch contour by conventional vs. proposed method

6. 결론

본 논문에서는 화상통신 및 인터넷폰 등의 용도로 많이 사용되는 G.723.1의 처리시간 개선을 위해서 부호화 과정중 피치 추정 알고리즘의 처리속도 단축을 제안하였다. 표준으로 사용되고 있는 시간영역 교차상관법이 과다한 곱셈연산을 수행함으로 인해 처리시간이 오래 걸린다는 단점을 보완하기 위하여 개선된 혼성영역 교차상관법에 의해 피치를 추정하도록 개선 하였다. 개선된 방법은 시간영역상의 교차상관값이 주파수영역상에서 켈레복소수 곱셈에 의하여 계산될 수 있다는 점과 처리 영역 변환 과정에 사용되는 FFT와

IFFT의 비트-재정렬 과정을 생략함으로써 시간영역 교차상관법에 비해 처리시간을 67.37% 정도로 단축하면서도 피치추정의 성능은 그대로 유지됨을 알 수 있었다.

**참 고 문 헌**

- [1] ITU-T Recommendation G.723.1, 2006. 5.
- [2] Embree, Paul M. & Bruce Kimble, C Language Algorithms for Digital Signal Processing, Prentice-Hall, 1991.
- [3] Wangrae Jo, Jongkuk Kim, Myungjin Bae, "A Study on Pitch Detection in Time-Frequency Hybrid Domain," Springer-Verlag, Lecture Notes in Computer Science, Vol.-LNCS3406, pp.437-440, February 2005.
- [4] 민소연, 정현욱, 배명진, "CELP 보코더에서 델타 피치 검색 방법 개선에 대한 연구," 2004년도 대한전자공학회 하계종합학술대회 논문집, Vol.27, No.1(s), pp.1181-1184, 2004년 6월.
- [5] 정익주, "실시간 디지털 신호처리," 생능출판사, 2006.
- [6] 미카미나오키(송봉길 역), C언어에 의한 디지털 신호처리 입문, 성안당, 2006.



**배 명 진 (裵 明 振)**

1986년~1992년 : 호서대학교 전자공학과 조교수  
1992년~현재 : 숭실대학교 정보통신공학과 교수

**저 자 소 개**



**조 왕 래 (趙 旺 來)**

1996년 2월 : 숭실대학교 정보통신공학과 (공학사)  
1998년 2월 : 숭실대학교 전기공학과(공학석사)  
2009년 2월 : 숭실대학교 정보통신공학과 (공학박사)  
2003년 ~ 현재 : 디비정보통신 책임연구원



**최 성 영 (崔 盛 永)**

1980년 2월 : 울산공과대학 전자공학과 (공학사)  
1996년 2월 : 숭실대학교 전산공학과(공학석사)  
2004년 8월 : 숭실대학교 전자공학과(공학박사)  
1990년~2006년 : 서울정보기능대학 통신전자과 부교수  
2006년 ~ 현재 : 한국폴리텍II대학 전자과 교수/산학협력단장