

## 보코더에서 프레임별 에너지 보상에 의한 피치검색 성능 개선에 관한 연구

백금란<sup>1</sup>, 민소연<sup>2</sup>, 배명진<sup>1\*</sup>

<sup>1</sup>송실대학교 정보통신학과, <sup>2</sup>서일대학교 정보통신과

## An Algorithm on Improving a Pitch Searching by Energy Compensation in a Frame for Vocoder

Geum-Ran Baek<sup>1</sup>, So-Yeon Min<sup>2</sup> and Myung-Jin Bae<sup>1\*</sup>

<sup>1</sup>Dept. of Information and Telecommunication, Soongsil University

<sup>2</sup>Dept. of Information&Communication Seoil University

**요 약** 보코더 과정 중 피치를 분석하는 과정은 코드북 분석과 더불어 부호화기 성능에 큰 영향을 미치는 부분이다. 피치를 검색하기 위해 먼저 신호의 주기성을 강조해야 하는데 일반적으로 자기상관법의 일종인 듀얼 펄스 알고리즘 기법을 이용하여 주기성을 강조한 후 피치를 검색하는 방법을 많이 사용한다. 이 방법은 한 프레임 내에서 두 개의 펄스 간격을 변경시키면서 음성신호와의 상관관계 값을 구하여 상관관계가 가장 높을 때의 펄스 간격을 찾는다. 또한 상관관계 값이 가장 높을 때의 주기가 가장 뚜렷한 반복 구간인데 이것을 피치 주기라 한다. 이때 반주기, 배주기 및 세배주기가 주된 주기로 찾아지는 경우에는 이 간격을 피치 주기라 할 수 없어 이를 해결하기 위한 여러 가지 방법들이 제안되어 있다. 본 논문에서는 피치를 검색하기 전에 프레임내의 전체 에너지변화 비율을 추정하여 신호의 에너지 레벨을 미리 보상해 준 후 피치를 검색하는 방법을 제안한다. 이 방법을 적용하면 피치검색 시간을 단축할 수 있고, 피치 검색의 정확도를 높일 수 있어 전반적인 피치 검색에 관한 성능이 개선된다.

**Abstract** It is important to search a pitch for vocoder. The major drawback to vocoders is their large computational requirements in searching a pitch and a codebook. In this paper, a simple method is proposed to improve the pitch searching process in the pitch filter almost without degradation of quality. The period of speech signal is emphasized by using Dual Pulse technique, the same type of autocorrelation method, in pitch search. Sometimes the incorrect pitch can be obtained by halving, doubling and trifling, To solve it, before searching a pitch, we estimate energy rate in a frame and compensate envelop of signal with it. By using the proposed algorithm in pitch search, its required computation are reduced and searching pitch is improved

**Key Words** : Pitch search, Autocorrelation, Dual pulse technique

### 1. 서론

음성신호를 메모리에 저장하거나 전송하기 위한 음성 부호화법에는 크게 파형부호화법, 신호원부호화법, 혼성 부호화법등 세 가지로 나눌 수 있다. 파형부호화법은 음성의 성분 분리 없이 파형 자체의 잉여 성분만을 제거한 후 부호화하여 전송하고 다시 합성하는 방식으로 고음질

과 화자의 개성이 유지되는 반면 전송파형을 유지하기 위한 데이터 량이 많기 때문에 전송률이 높으며 대용량의 메모리가 필요하게 된다[1]. 이에 비해 신호원 부호화법은 음성의 발생모델에 근거하여 음성 신호의 여기 성분과 여파기 성분을 분석하여 각각을 독립적으로 분리시켜 부호화하는 방법을 사용하기 때문에 전송 대역폭이 작고 메모리 사용량이 작다. 그러나 분석 및 합성 시에

\*Corresponding Author : Myung-Jin Bae

Tel: +82-2-824-0902 email: mjbae@ssu.ac.kr

접수일 12년 05월 29일

수정일 (1차 12년 07월 03일, 2차 12년 07월 11일)

게재확정일 12년 07월 12일

오차가 누적되어 명료성이 떨어진다는 단점을 갖고 있다 [1]. 혼성 부호화법은 고음질의 화자의 개성을 유지시켜 주는 파형 부호화법과 메모리 효율성이 높은 신호원 부호화법을 결합시킨 방법이다. 즉 포먼트 정보는 선형 예측 부호화법으로 부호화하고 나머지 잔여 신호를 어떻게 부호화하느냐에 따라 RELP, VELP, MPLPC, CELP 등이 제안되어져 있다[1,5].

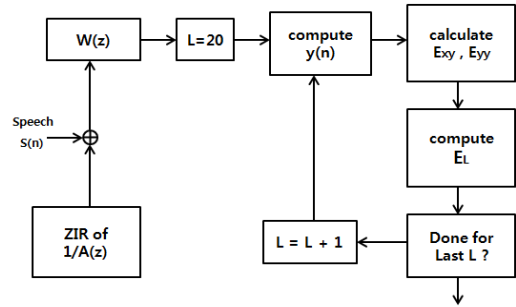
음성이 선형예측을 통해 얻어 질 수 있다는 가정으로 개발된 CELP(Code Excitation Linear Prediction)계열의 음성부호화기들은 음성을 부호화하는 데 있어서 선형예측 계수와 여기신호(excitation signal)로 이루어진다. 일반적으로 선형 예측 계수는 LSP를 이용하여 부호화되며 여기신호는 몇 개의 코드북(codebook)을 이용하여 부호화한다. CELP를 기반으로 발전한 부호화 기법으로는 ACELP, CS-CELP 등이 있다. 또한 CELP계열의 압축 방식은 낮은 전송률에서 좋은 통화 품질을 제공해 준다 [1,2,7]. 음성 부호화 기술로 현재 거의 주로 사용되는 G.723.1, G.729, AMR, iLBC EVRC 등이 CELP 기반의 음성 부호화기이다. 현재 ITU-T에서는 PCS, IMT-2000등에서 사용할 수 있는 8kbps 음성 부호화기에 대한 표준화 작업으로 1996년에 CS-CELP를 G.729로 인터넷 및 화상 통신용 음성 부호화기로 ACELP/ML-MLQ의 5.3/6.3kbps 이중 속도 부호화 방식을 G.723.1로 선정하였다[9].

앞에서 살펴 본 CELP보코더의 과정 중 피치(pitch)를 검색하는 과정은 코드북 검색과 함께 매우 중요한 과정이다. CELP부호화법에서는 피치필터를 적용하여 음성신호의 피치 주기 성분을 부호화하고 있다. 이러한 피치필터에 주로 적용되고 있는 피치검색법은 피치지연에 따른 상관관계법이다. 상관관계 피치검색법은 피치가 존재하는 모든 피치지연에 대해 원래음성과 합성된 음성의 상관관계를 검색하고, 최상의 상관관계를 갖는 피치필터의 피치지연과 이득을 결정하게 된다[5].

본 논문에서는 프레임 별 에너지 비를 이용하여 이득을 미리 보상하여 최적의 피치지연을 구하는 방법을 제안한다. 이 방법으로 피치를 검색하는 시간을 단축시킬 뿐만 아니라 피치 검색의 정확도를 높여주어 전송하고자 하는 음성신호의 파라미터들을 코딩하는 것도 용이하게 해준다. 먼저 2장에서 CELP계열의 보코더에서의 피치 검색 알고리즘을 살펴보고 3장에서 본 논문에서 제안하는 프레임 별 에너지 비 보상 알고리즘을 설명한다. 4장과 5장에서는 실험 및 결과를 살펴보고 결론을 맺는다.

## 2. CELP계열 보코더에서의 피치검색

보코더 부호화시에 수행되는 피치 검색과정은 음성신호의 자기 상관관계에 해당하는 피치 주기 정보를 얻는 과정이다. 스펙트럼분석이 개회로 구조로 수행되어지는 것에 비해 피치분석은 폐회로 구조에 의해 수행되어져야 한다. 즉, 폐회로 구조에 의해 수행함으로써 피치 지연조건을 최적으로 만족하는 값을 반복적인 비교를 통해 결정할 수 있게 된다. 피치 이득은 이때 얻어진 피치지연을 이용하여 상대적인 이득 값을 갖도록 양자화하게 된다. 이러한 피치 검색과정은 코드북 검색과 함께 CELP 부호화기의 계산 양에 크게 영향을 미치는 중요한 부분이다[5].



[그림 1] 피치 검색을 위한 구현 과정의 예  
[Fig. 1] An example of implementation flow for pitch search.

그림 1은 전형적인 피치검색 과정 중 한 과정만을 보여 주고 있다. 피치검색은 합성에 의한 분석을 기초로 하고 피치 예측 필터를 위한 입력신호와 합성된 신호사이의 가중화된 에러를 최소로 하는 지연 값(L)과 이득 값(b)를 선택하는 과정이다. 그림1에서 ZIR(Zero Input Response)은 영(zero) 입력 응답이고 피치검색을 더욱 효과적으로 하기위해 가중여파기,  $W(z)$ 를 사용하는데 식(1)과 같다. 원래 신호와 합성신호에 다음의 필터를 씌워 에러를 구한다.

$$W(z) = \frac{A(z)}{A(z/\alpha)} \quad (1)$$

여기서  $A(z)$ 는 포먼트(formant) 예측 오차 필터이고  $\alpha$ 는 인지 가중화된 파라미터이다. 또한 피치 합성 필터는 식(2)가 된다.

$$\frac{1}{P(z)} = \frac{1}{1 - bz^{-L}} \quad (2)$$

가장 최적의  $L$ 과  $b$ 를 구하는 과정은 식(3)의 MSE(Mean Square Error)를 최소화하는  $L$ 과  $b$ 을 구하는 것과 같다.

$$MSE = \frac{1}{L_p} \sum_{n=0}^{L_p-1} [x(n) - by_L(n)]^2 \quad (3)$$

$$= \frac{1}{L_p} \sum_{n=0}^{L_p-1} [x(n) - by_L(n-L)]^2$$

여기서  $x(n)$ 과  $y(n)$ 은 각각 입력신호와 인지 가중화된 합성 신호이고  $L_p$ 는 피치분석 프레임 길이이다. 이것은 식(4)의 최대값을 구하는 것과 같다.

$$E_L = \frac{(E_{xy})^2}{E_{yy}} \quad (4)$$

여기서  $E_{xy} = \sum_{n=0}^{L_p-1} x(n)y_L(n)$

$$E_{yy} = \sum_{n=0}^{L_p-1} y_L(n)y_L(n) \text{ 이다.}$$

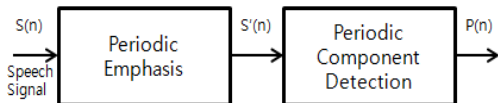
주어진  $L$  값에 대해 최적의  $b$ 는 식(5)에서 구한다.

$$b_L = \frac{E_{xy}}{E_{yy}} \quad (5)$$

이러한 과정은 각각의 모든  $L$ 값에 대해 반복적으로 수행된다[9].

### 3. 제안하는 알고리즘

일반적으로 피치를 검색하기 전에 그림2와 같이 신호의 주기성을 먼저 강조한 다음 그 강조된 신호에서 두드러진 반복 구간을 찾는다.



[그림 2] 피치 검색 과정  
[Fig. 2] Pitch search process

자기상관법에 속하는 듀얼펄스(dual pulse) 기법은 두 개의 임펄스 신호 쌍과 원래신호의 상관관계를 구하여 가장 두드러진 반복구간을 찾는다. 그러므로 신호의 레벨은 중요하지 않고 간격이 중요하다. 즉 한 프레임 내에서 식(6)과 같은 전달특성을 갖는 두 개의 펄스와 원래 음성 신호와의 가장 높은 상관관계 값을 구하여 주된 주기를 찾는 방법이다[8.10].

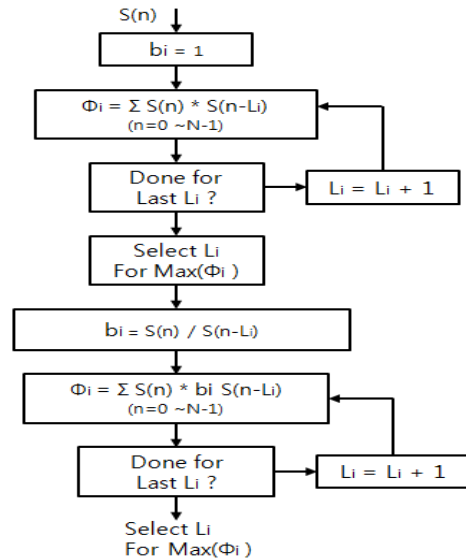
$$H(z) = 1 - bz^{-L} \quad (6)$$

여기서  $b$ 은 필터의 이득이며  $L$ 은 두 펄스간의 간격이다.

$$\Phi_i = \sum_{n=0}^{N-1} s(n) * b_i s(n - L_i) \quad (7)$$

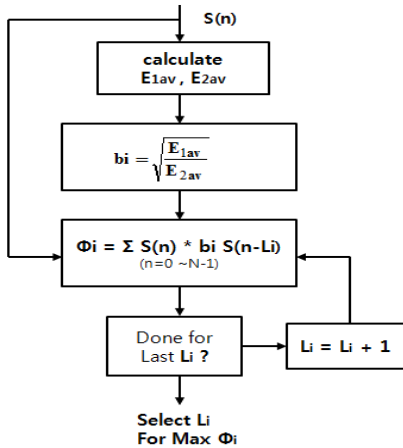
(단,  $N$ 은 프레임사이즈)

식(7)을 이용해서 프레임 내에서 샘플을 증가시켜가며 원래 신호와 두 펄스 간격만큼 지연된 신호 사이의 상관관계를 계산한다. 여기서  $b_i, L_i$ 는 각각  $i$ 번째 프레임의 이득과 펄스 간격이다. 이 식에서 얻어진 파형에서의 두드러진 피크와 피크 간격 중에 첫 번째로 찾아지는 가장 큰  $L_i$ 값을 피치 주기라 한다. 그러나 이 경우에 찾아진  $L_i$ 값이 반드시 피치라는 보장이 없다. 이유는 이 간격이 피치 주기가 아닌 반주기, 배주기 및 세배주기일 수 있다.



[그림 3] 듀얼펄스 기법에 의한 피치 검색 과정  
[Fig. 3] Pitch search by dual pulse

반주기가 배주기가 피치 주기로 찾아지는 것을 막기 위해 그림 3에서 나타났듯이 처음에  $b_i$  값을 1(one)로 놓고  $L_i$ 값을 증가 시켜가며 원래 신호와  $H(z)$ 의 상관관계를 구한다. 즉 시간영역에서의 식(7)의 값을 구한다. 이때 상관관계 값이 가장 높을 때의 신호  $S(n)$ 과  $L_i$ 만큼 지연된 신호  $S(n - L_i)$ 의 비를 구하여  $b_i$  값으로 정한다. 두 번째 단계로 이렇게 얻어진  $b_i$  값을 다시  $H(z)$ 에 대입하여  $L_i$  값을 증가 시켜가며 원래 신호와  $H(z)$ 의 상관관계를 구한다. 즉 다시 식(7)의 값을 구한다. 이런 과정들을 통해 얻어진 상관관계 값인  $\Phi_i$ 가 가장 높을 때의  $L_i$  값을 피치 주기로 한다. 이렇게 하면 반주기, 배주기 및 세배주기가 피치 주기로 찾아지는 것을 줄여 준다. 그러나 프레임 마다 이런 과정을 반복하다 보면 원하는 값을 얻기 위해서는 많은 계산 량이 필요하다는 것을 알 수 있다.



[그림 4] 제안한 알고리즘의 블록도  
[Fig. 4] Block diagram of the proposed algorithm

본 논문에서는 그림 4에서 나타낸 것처럼 한 프레임 내에서 신호 파형의 평균진폭 변화율을 구하여 이를  $b_i$  값으로 놓는다. 즉, 피치주기를 검색하기 전에 미리  $b_i$ 를 구하여 변화율만큼을 미리 진폭에 보상하여 준 후 피치 주기를 구하는 방법을 제안한다.

$$b_i = \sqrt{\frac{E_{1av}}{E_{2av}}} \quad (8)$$

( 단,  $E_{av} = \frac{1}{M} \sum_{n=1}^M S^2(n)$  , M은 샘플수 )

먼저 피치를 검색하기 전에 프레임 앞부분(약10msec)의 평균에너지인  $E_{1av}$ 과 뒷부분의 평균에너지(약 10msec)인  $E_{2av}$ 를 각각 구한다. 이렇게 구한 프레임 앞뒤에서의 에너지 값을 이용하여 식(8)에서처럼 에너지비율인  $i$ 번째 프레임의  $b_i$ 값을 정한다. 이런 과정이 가능한 이유는 그림3에서 얻어진  $b_i$  값은 신호의 진폭의 변화되는 정도를 나타내는 것이므로 처음부터 진폭의 변화율인 에너지 비를 구하여 이를  $b_i$ 로 정할 수 있다. 이 알고리즘은 피치 검색을 위해 상관관계 값을 구하기 위해 여러 번의 많은 계산을 수행하지 않고 피치검색 전에 에너지 비율을 구해 진폭을 미리 보상한 후 피치를 검색하므로 검색시간을 단축할 수 있고 피치검색의 정확도를 좀 더 높일 수 있다.

#### 4. 실험 및 결과

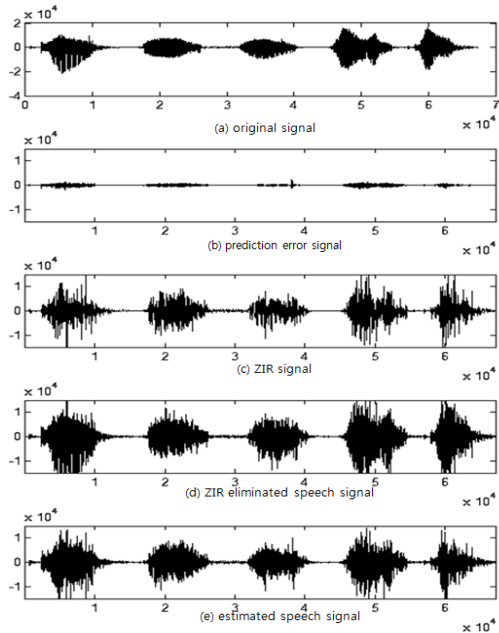
본 논문에서 제안한 방법을 실험하기 위해 시뮬레이션 장비로는 워크스테이션 Z400을 사용하였고 음성 시료는 8kHz로 샘플링(Sampling)했으며 16bit A/D 컨버터로 디지털 신호로 변환된 20대 남/녀의 음성을 이용하였다. 또한 제안한 알고리즘의 성능을 평가하기 위해 피치 검색 과정을 Visual C++과 MatLab 프로그램으로 구현하여 기존의 CELP 보코더에서의 피치 검출 방법과 본 논문에서 제안한 방법을 비교하였다.

프레임의 간격은 160 샘플로 정하였고 식(8)의 값 즉,  $i$  번째 프레임의 이득인  $b_i$ 를 구하기 위해 프레임의 앞과 뒤 파형의 에너지 비를 구할 때는 각각의 프레임 시작과 끝부분으로부터 일정 간격 만큼의 에너지를 구하여 계산하였다. 그 일정 간격은 10msec, 9msec, 8msec등 줄여가며 실험하였다.

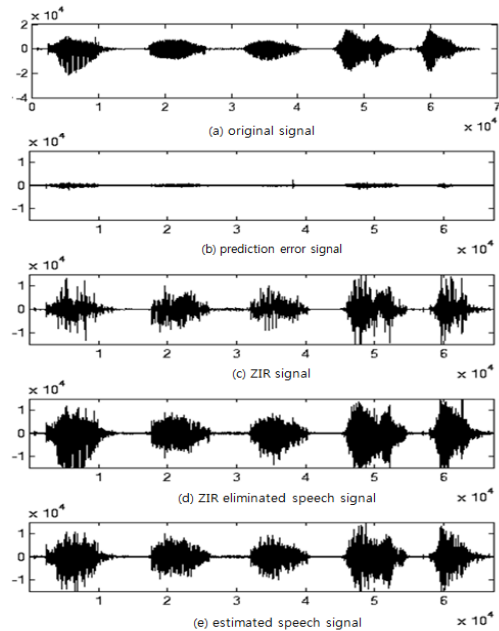
그림 5는 QCELP 알고리즘에 의한 원래음성으로 부터 최적의  $L$ 과  $b$ 를 추출하여 다시 합성 신호를 만들어 내는 과정을 그림으로 나타낸 것이다. (a)는 원래 신호 파형, (b)는 주기적 에러신호, (c)는 ZIR 신호 파형, (d)는 ZIR가 제거된 신호파형 그리고 (e)는 최적의  $L$ 과  $b$ 에 의해서 합성된 신호이다.

그림 6은 본 논문에서 제안한 방법에 의해 합성 신호를 얻는 과정을 그림으로 나타낸 것이다. 원래 음성과 똑 같지는 않지만 어느 정도 유사성을 가지고 있으며 음성 신호의 모양은 다르지만 실제로 들었을 때는 QCELP방법으로 합성된 음성과 비교했을 때 큰 차이를 느끼지 못

할 정도로 음질이 비슷했다. 오히려 ZIR가 제거된 신호의 음질은 본 논문에서 제안한 방법이 더 좋았다.



[그림 5] QCELP 보코더에서의 원음성과 합성음성  
[Fig. 5] Original speech and synthesized speech in QCELP vocoder



[그림 6] 제안한 알고리즘의 실험 결과  
[Fig. 6] Original voice and synthesized voice in proposed algorithm

## 5. 결론

자기상관함수 방법에 속하는 듀얼펄스 기법은 주기성을 강조한 후 피치를 검색하는 방법이다. 한 프레임 내에서 두 개의 펄스 간격을 변경시키면서 상관관계 값을 구하여 주기가 가장 두드러지는 반복 구간을 찾아 피치를 검색한다. 이때 반주기, 배주기 및 세배주기가 발생하는 경우에는 찾아진 간격을 피치라 할 수 없어 이를 해결하기 위한 여러 가지 방법들이 제안되어있다.

본 논문에서는 피치를 검색하기 전에 프레임내의 전체 에너지변화율을 추정하여 피치 검색 전에 신호의 에너지레벨을 보상 하여 피치검색 성능을 향상시키는 방법을 제안한다. 이로서 반주기, 배주기 및 세배주기 줄일 수 있어 좀 더 정확한 피치를 검색할 수 있으며 전송하고자 하는 음성신호의 파라미터들을 코딩하는 경우에도 용이하게 된다. 향후 제안한 알고리즘을 다른 보코더에도 적용하여 계산 양의 감소 정도를 측정하고 객관적인 음질을 평가를 통해 제안한 알고리즘의 성능을 평가할 것이다.

## References

- [1] M. J. Bae, *Digital Speech Coding*, p17-29, p30-48, pp49-68 Dongyoung, 1996.
- [2] J. S. Han, *Speech Processing*, Osung Media, pp.5- pp7, pp103. 2003년.
- [3] M. J. Bae, S. H. Lee, *Digital Speech Analysis*, p123- Dongyoung, 1996.
- [4] L. R. Rabiner, R. W. Schafer, *Digital Processing of Speech Signal*, pp.131-149.
- [5] M. J. Bae, "On a Performance Comparison of Pitch Search Algorithms by using a Correlation Properties for the CELP Vocoder", In ASK, the proceeding Vol. 11, pp. 165-171, 2004.
- [6] J. K. Kim and M. J. Bae, "A study of pitch extraction method by using harmonics peak-fitting in speech spectrum", In ASK, Proceedings of ICSP 2001, Vol, 1, pp. 617-622, 2001.
- [7] A. M. Kondoz, *Digital Speech coding for low bit rate communications systems*, John Wiley & Sons, new York, 2004.
- [8] JooHun Lee, MyungJin Bae, SouGil Ann, "A Fast Pitch Searching Algorithm Using Correlation Characteristics in CELP Vocoder", The Journal of ASK, Vol. 13. No. 2E, 1994.
- [9] Wang-Rae Jo, Seong-Young Choi, MyungJin Bae, "A

Study on Pitch Search Time Reduction of G.723.1 Vocoder by Improved Hybrid Domain Cross-correlation" Trans.KIEE.Vol.59,No.12,DEC,2010.

- [10] JaeOk Bae, KiChun Han, JinJong Cha and MyungJin Bae, "On a Fast Searching of the Random Codebook using the Searching Technique of the Regular Pulse Excitation," International Conference on Speech Processing, Proceedings of ICSP'97, pp.281-286, August 26-28, 1997.

**배 명 진(Myung-Jin Bae)**

[정회원]



- 1986년 3월 ~ 1992년 2월 : 호서대학교 조교수
- 1992년 3월 ~ 현재 : 송실대학교 정보통신공학과 정교수

**백 금 란(Geum-Ran Baek)**

[정회원]



- 1994년 2월 : 송실대학교 전자공학과 (공학사)
- 1999년 2월 : KAIST 전기 및 전자공학과(공학석사)
- 2012년 2월 : 송실대학교 정보통신공학과 (박사수료)
- 2009년 3월 ~ 현재 : 송실대학교 전산원 교수

<관심분야>

정보통신, 신호 및 신호처리, 음성신호처리

<관심분야>  
정보통신

**민 소 연(So-Yeon Min)**

[중신회원]



- 1993년 2월 : 송실대학교 전자공학과 (공학사)
- 1999년 2월 : 송실대학교 전자공학과(공학석사)
- 2003년 2월 : 송실대학교 전자공학과(공학박사)
- 2005년 3월 ~ 현재 : 서일대학교 정보통신과 교수

<관심분야>

음성신호처리, 음성코딩, 멀티미디어통신