

QMF에 의한 음성스펙트럼의 평탄화 알고리즘

민소연^{1*}

The Flattening Algorithm of Speech Spectrum by Quadrature Mirror Filter

So-Yeon Min^{1*}

요 약 음성신호의 프리-엠퍼시스 과정은 고주파 영역의 약화된 성분을 보상하기 위해 사용되어진다. 프리-엠퍼시스 필터의 일반적인 형태는 $y(n) = s(n) - A \cdot s(n-1)$ 이고, A 값은 유성음의 경우 0.9 ~ 1.0 사이의 값이다. 그리고, A 값은 프리-엠퍼시스의 기울기 값을 반영하고 기존의 방법에서는 R(1)/R(0)를 사용한다. 본 논문에서는 성문특성으로 인해 고주파 특성이 약화되는 것을 보상하기 위하여 새로운 평탄화 기법을 제안한다. 우선 신호 왜곡의 최소화를 위하여 QMF를 사용하였다. QMF를 사용한 후, 각 프레임별 자기상관계수를 사용하여 평탄화 과정을 수행하였다. 실험결과에서는 제안한 방법이 자기상관 방법보다 약화된 고주파 성분을 효과적으로 보상하는 평탄화 특성이 우수한 것으로 나타났다. 그러므로 평탄화 알고리즘은 음성 인식, 음성 분석 및 합성 등과 같은 음성 신호 처리 분야에 광범위하게 적용되어질 수 있다.

Abstract Pre-emphasizing the speech compensates for falloff at high frequencies. The most common form of pre-emphasis is $y(n) = s(n) - A \cdot s(n-1)$, where A typically lies between 0.9 and 1.0 in voiced signal. And, this value reflects the degree of pre-emphasis and equals R(1)/R(0) in conventional method. This paper proposes a new flattening method to compensate the weakened high frequency components that occur by vocal cord characteristic. We used QMF(Quadrature Mirror Filter) to minimize the output signal distortion. After using the QMF to compensate high frequency components, flattening process is followed by R(1)/R(0) at each frame. Experimental results show that the proposed method flattened the weakened high frequency components effectively than auto correlation method. Therefore, the flattening algorithm will apply in speech signal processing like speech recognition, speech analysis and synthesis.

Key Words : QMF(Quadrature Mirror Filter), 프리-엠퍼시스 필터, 자기상관법

1. 서 론

음성신호는 소리와 반복으로 이루어지며, 소리와 그 사이의 변이는 정보에 대한 기호적인 표현으로 나타난다. 소리에 대한 배열은 언어의 규칙에 의해 결정되며, 이 규칙과 인간의 통신에 있어서 의미에 대한 연구는 언어학의 영역이며, 음성의 소리를 연구하고 분류하는 것은 음성학 분야이다[1]. 음성신호의 구조에 대한 연구는 음성

정보를 추출하거나 강조할 수가 있다. 따라서 음성신호의 생성에 대한 수학적 모델은 음성을 처리하는데 있어서 매우 중요하다.

음성신호의 구조에 대한 연구는 음성정보를 추출하거나 강조할 수가 있다. 따라서 음성신호의 생성에 대한 수학적 모델은 음성을 처리하는데 있어서 매우 중요하다. 음성에 의한 의사전달은 말하는 이가 전달 하고자하는 생각에서 출발하여 이 생각을 일련의 신경과정과 근육의 움직임을 통해 음향 압력 파로 듣는 이의 청각 시스템에 따라 신경신호로 바꾸어 처리하게 된다. 이러한 과정을 음성 발생과정과 음성인식과정에서 구분한다. 음성신호는 성문특성으로 인해 고주파 특성이 약화되는 경향이

본 연구는 서울시 산학협력사업으로 구축된 서울 미래형콘텐츠컨버전스 클러스터 지원으로 수행되었습니다.

¹서일대학 정보통신과

*교신저자: 민소연(symin@seoil.ac.kr)

있으며 이를 보상하기 위해 프리-엠퍼시스 필터가 사용되어진다.

간단히 수식으로 표현하면 시간영역에서 $y(n) = s(n) - As(n-1)$ 와 같이 차분 방정식으로 나타낼 수가 있다[3,4]. 여기서 A값은 보통 0.9에서 1사이의 값을 주로 사용하여 고주파 특성이 약화되는 경향을 보상한다. 그러나 음성의 스펙트럼에서 전체 주파수 중에 양 끝부분의 주파수영역은 잘 보상되나 중간영역에 위치한 주파수 부분은 정확히 보상되지 않는다. 본 논문에서는 QMF 필터를 이용하여 입력신호를 고주파와 저주파의 2개의 대역으로 분할하여 각 밴드에 프리-엠퍼시스 필터를 적용하여 약화된 특성을 정확히 보상한다. 본 논문의 전체적인 구성은 2장에서, QMF 필터에 대한 이론을 간단히 설명하고 3장에서는 프리-엠퍼시스 필터 특징에 대해 알아보도록 한다. 4장에서는 본 논문에서 제안한 QMF 필터에 의해서 음성스펙트럼의 기울기를 보상하는 방법을 설명하고 5장에서는 실험 및 결과, 6장에서 결론을 맺는다.

2. QMF(Quadrature Mirror Filter) 분석

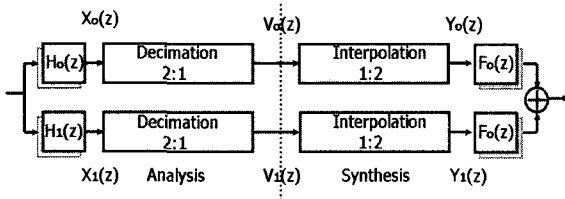


그림 1. QMF 서브밴드 필터의 기본구조

QMF[5] 필터는 그림 1과 같이 기본적으로 분해 필터(Analysis filter)와 합성 필터(Synthesis filter)로 이루어지며 입력신호는 필터에서 저주파와 고주파, 두 부분으로 나누어진다[5]. 분해 필터는 입력신호를 식 (2.1)과 같이 고주파와 저주파, 2개 대역으로 분할하고 각 분할된 신호를 2로써 다운 샘플링(Down-sampling) 한다. 합성 필터는 식 (2.2)와 같이 분해 필터에서 분해된 각 밴드의 신호들을 2로써 업 샘플링(Up-sampling)하고 합성 필터를 통과시킨 후 합성한다.

일반적으로 분해 필터를 대역 필터로 사용하여 다운 샘플링하면 간섭현상(aliasing)이 발생한다. 이를 제거하고 신호 성분만을 위해서 분해 필터와 합성 필터를 공역 직교 필터로써 구성한다[5][7].

$$X_k(Z) = H_k(Z) * X(Z) \quad K=0,1 \quad (2.1)$$

$$V_k(Z) = \frac{1}{2} [X_k(Z^{1/2}) * X_k(-Z^{1/2})]$$

$$Y_k(Z) = V_k(Z^2) = \frac{1}{2} [H_k(Z)X(Z) + H_k(-Z)X(-Z)]$$

$$X'(Z) = [F_0(Z)Y_0(Z) - F_1(Z)Y_1(Z)]$$

$$X'(Z) = \frac{1}{2} [H_0(Z)F_0(Z) + H_1(Z)F_1(Z)]X(Z) + \frac{1}{2} [H_0(-Z)F_0(Z) + H_1(-Z)F_1(Z)]X(-Z)$$

$$F_0(Z) = H_1(-Z), F_1(Z) = -H_0(-Z)$$

$$X'(Z) = \frac{1}{2} [H_0(Z)H_1(Z) - H_1(Z)H_0(-Z)]X(Z) \quad (2.2)$$

$$|H_0(Z)|^2 + |H_1(Z)|^2 = 1$$

여기서 Ho(z), H1(z)은 저주파 및 고주파 대역통과 분해 필터이고, Fo(z), F1(z)은 저주파 및 고주파 대역통과 합성 필터이다[5][7], 즉, 식(2.1)과 (2.2)는 고주파 영역에서 약화된 음성신호를 보상하기 위하여 통과하는 QMF 필터의 분해 필터와 합성 필터에 대한 설명을 나타내고 있다.

3. 프리-엠퍼시스 필터의 특징

음성신호에서는 음성의 의미를 내포하고 있는 음향학적 특성은 저주파 영역에 나타나고 있지만 화자의 특징은 고주파 영역에 많이 분포되어 있다. 따라서 화자 인식의 전 처리 단계에서 고주파 영역을 강조시키기 위해 사용되는 프리-엠퍼시스 필터의 역할이 요구된다. 음성신호처리 분야에서 스펙트럼 경사를 평탄화해 줌으로써 신호의 동적 범위를 억제하는 프리-엠퍼시스 과정은 신호 대 잡음비(SNR[7]:Signal to Noise Ratio)를 높이는데 유효한 것으로 알려져 있다. 이 방법은 일반적으로 A/D 변환을 위한 저역 통과 필터링에 앞서서 20dB/decade 정도의 고주파 영역을 강조하는 역할을 한다.

이 방법은 또한 A/D 변환 다음에도 수행되어질 수 있는데, 차분 방정식이나 식 (3.1)과 같은 1차 디지털 필터링을 통하여 구현된다.

$$H(z) = 1 - az^{-1} \quad (3.1)$$

여기에서 a 값은 대략 1에 가까운 값을 갖는다. 신호 대 잡음비(SNR)를 가능한 높이기 위해서는 A/D 변환을 하기 앞서 프리-엠퍼시스를 해야 한다. 그리고 원 신호의 스펙트럼 경사를 복구하기 위한 과정을 디엠퍼시스(de-emphasis)라고 한다[6,7,8,9].

4. 제안한 주파수 영역에서의 기울기 측정

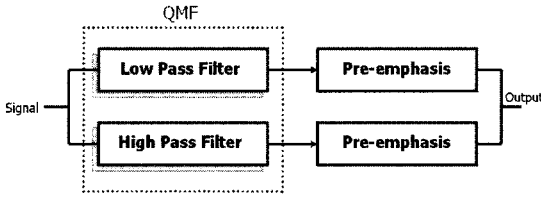


그림 2. 제안한 방법의 블록도

그림 2는 본 논문에서 주파수 영역의 모든 성분들을 평탄화 하기 위해 제안한 블록도이다. 입력 음성신호를 받고 각 프레임마다 해밍 윈도우를 취한다. 윈도우를 취한 데이터를 이용해 QMF(Quadrature Mirror Filter)를 통과시킨다. 음성신호는 저주파와 고주파, 두 부분으로 나뉘지며 저주파 부분의 신호를 프리-엠퍼시스 과정을 적용한다. 프리-엠퍼시스 필터는 스펙트럼 상에서 저주파 영역을 감쇄시키고 고주파 영역을 강조시킴으로써 신호를 평탄화 시키는 특성을 가지고 있다.

$y(n) = s(n) - As(n-1)$ 에서 A 값은 대체적으로 고정된 값을 적용하나 이 방법은 고주파 영역의 에너지 성분 모두를 강조시키기 때문에 인지에 중요한 특성을 보이고 있는 공진 봉우리 성분뿐만 아니라 인식에 불필요한 성분들도 강조하게 되는 결과를 초래하게 된다. 또한 프리-엠퍼시스 필터에 통과된 음성 스펙트럼이 완전히 평탄화 되지 못하기 때문에 인식시 저주파의 영향이 많이 지배하는 특징을 가지고 있다. 따라서 고주파 영역을 강조하더라도 성분특성을 이용하여 적응적으로 적용하고 스펙트럼을 평탄화 시키는 전 처리 단을 구성하고자 한다. 성분특성을 고려한 필터계수를 구하는 방법은 다음과 같다.

단 구간 자기상관 함수는 식 (4.1)로 표현 가능하다 [4,8,9].

$$\Phi_n(i, j) = \sum_{m=0}^{N-1-(i-j)} s_n(m) s_n(m+i-j), \quad 1 \leq i \leq p, 0 \leq j \leq p \quad (4.1)$$

$$R_n(j) = \sum_{m=0}^{N-1-j} s_n(m) s_n(m+j) \quad (4.2)$$

$$\sum_{j=1}^p a_j \Phi_n(i, j) = \Phi_n(i, 0), \quad \text{for } i = 1, \dots, p \quad (4.3)$$

자기 상관법(Auto-correlation Method)을 이용하여 식 (4.3)을 풀면 다음과 같이 표현된다.

$$\begin{vmatrix} R_n(0) & R_n(1) & \dots & R_n(p-1) \\ R_n(1) & \dots & \dots & R_n(p-2) \\ \vdots & \vdots & \vdots & \vdots \\ R_n(p-1) & \dots & \dots & R_n(0) \end{vmatrix} \begin{vmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{vmatrix} = \begin{vmatrix} R_n(1) \\ R_n(2) \\ \vdots \\ R_n(p) \end{vmatrix} \quad (4.4)$$

$p=1$ 에 대하여 위의 식을 정리하면 다음과 같은 식으로 표현 가능하다.

$$a_1 = \frac{R_n(1)}{R_n(0)} \quad (4.5)$$

이 계수 값을 $y(n) = s(n) - As(n-1)$ 에 적용하여 스펙트럼경사를 평탄화 시킨다. 이 과정을 QMF 필터로 인한 고주파 부분에도 반복한다. 보상된 고주파 부분과 저주파부분을 합성하여 신호로 복원한다[4]. 그림 2에서는 제안한 평탄화 알고리즘의 전체 블록도를 나타내었다. 즉, 2장에서 설명되어진 QMF를 사용하여 저주파 대역과 고주파 대역으로 분리하여 각 주파수 특성에 맞게 프리-엠퍼시스 과정을 거치게 되므로 보다 음성신호의 프레임 별 특성에 적응적인 처리가 가능하다.

5. 실험 및 결과

본 논문에서 모의실험을 위해 IBM PC에 마이크 입력이 가능한 A/D 변환기를 인터페이스 시키고 발생한 시료는 8kHz로 샘플링하고 16비트로 양자화된 아나운서 목소리를 사용하였다. 한 프레임의 길이는 256샘플이며 계수는 10차를 써서 특징벡터를 추출하였다. 그림 3은 유성음의 경우 제안한 방법을 이용한 결과를 스펙트럼으로 비교 분석한 것이다. (a)는 유성음의 음성신호 파형을 나타낸 것이며 (b)는 음성신호를 스펙트럼 분석한 것이다. (c)는 자기 상관함수를 이용하여 프리 엠퍼시스 필터의 계수 A 값을 적절히 조절하여 음성스펙트럼을 보상한 것을 보여 준다. (d)그래프를 보면 정확히 보상된 음성스펙트럼을 볼 수 있는데 이는 QMF 필터를 이용하여 저주파

와 고주파의 서브밴드로 나누어 각각 자기상관함수를 이용하여 각 밴드에 적용하여 보상된 것이다. 그림 3에서 (c)와 (d)그래프를 비교 분석해 보면 (c)는 한 프레임의 전체를 적절히 A값을 조절하여 보상하였으나 필터 성질을 보면 양 끝 부분의 주파수영역을 주로 보상하는 것을 볼 수 있다.

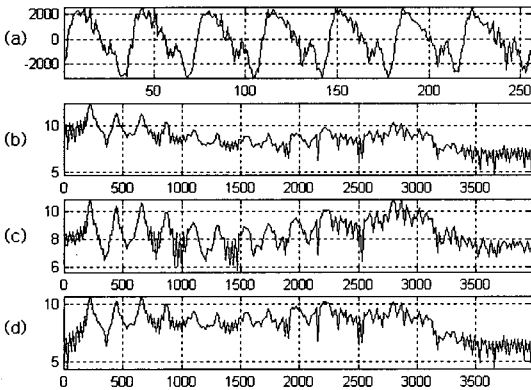


그림 3. 유성음의 스펙트럼 비교 분석
 (a) 음성신호
 (b) 음성신호의 스펙트럼 분석
 (c) 기울기 A값 조절을 통한 음성신호의 스펙트럼 분석
 (d) 제안한 방법을 이용한 스펙트럼분석

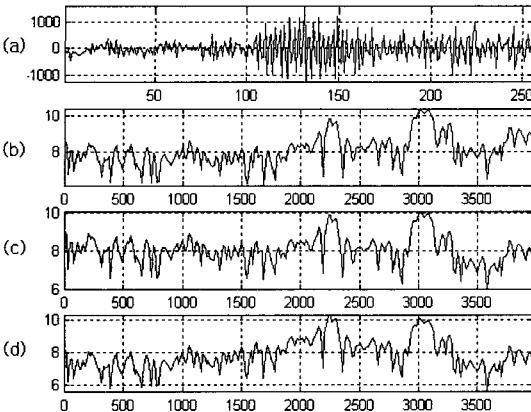


그림 4. 무성음의 스펙트럼 비교 분석
 (a) 음성신호
 (b) 음성신호의 스펙트럼 분석
 (c) 기울기 A값 조절을 통한 음성신호의 스펙트럼 분석
 (d) 제안한 방법을 이용한 스펙트럼분석

그러나 (d)는 서브밴드를 이용하여 중간대역의 주파수 까지 보상하여 좀 더 정확히 보상된 것을 알 수 있다. 그림 4는 유성음 프레임의 경우와 마찬가지로 무성음에 해당하는 프레임을 제안한 방법에 의해 음성스펙트럼을 보정한 후 그래프로 나타낸 그림이다. 그림 5,6은 주파수 영역에서의 스펙트럼 포락선을 좀 더 자세히 분석하기 위하여 LPC(Linear Predictive Coding) 분석 결과를 나타내었다. 그림 5의 (a)는 음성신호의 LPC 분석을 나타낸 그래프이며 (b)는 기울기 A값 조절을 통한 음성신호를 LPC로 분석한 것이다. (c)그래프는 QMF 필터를 통과한 후 프리-엠퍼시스를 거친 음성의 LPC 분석을 보여준다.

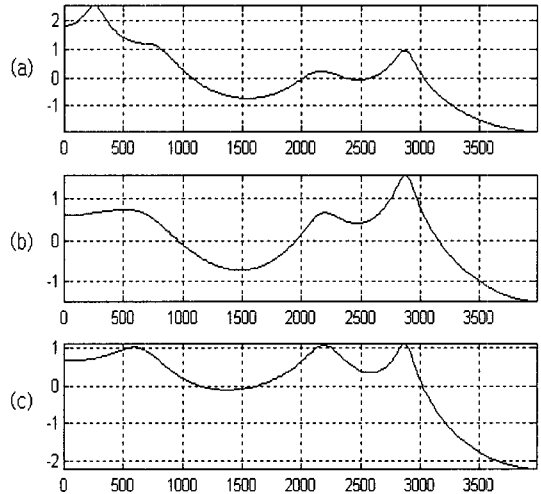


그림 5. 유성음의 LPC 비교 분석
 (a) 원음성의 LPC분석
 (b) 자기상관법을 이용한 LPC분석
 (c) 제안한 방법을 이용한 LPC분석

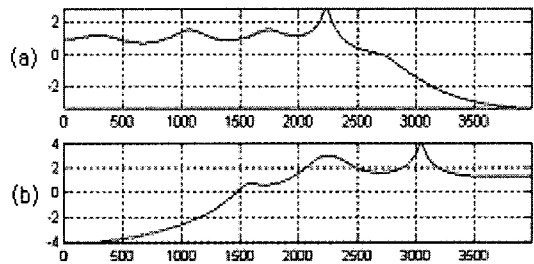


그림 6. 유성음 프레임의 저/고주파 분석
 (a) 제안한 방법의 저주파대역의 LPC 분석
 (b) 제안한 방법의 고주파대역의 LPC 분석

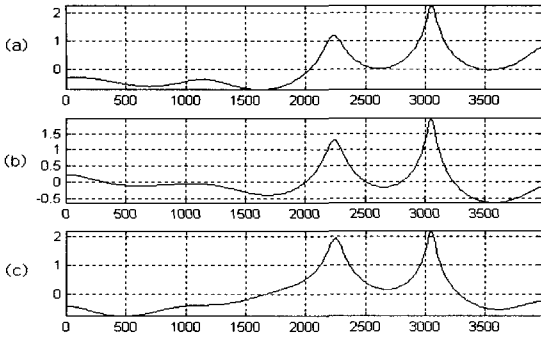


그림 7. 무성음의 LPC 비교 분석
 (a) 원음성의 LPC분석
 (b) 자기상관법을 이용한 LPC분석
 (c) 제안한 방법을 이용한 LPC분석

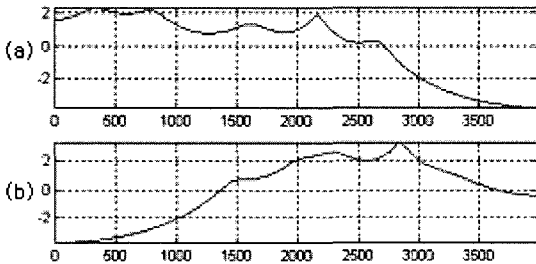


그림 8. 무성음 프레임의 저/고주파 분석
 (a) 제안한 방법의 저주파대역의 LPC 분석
 (b) 제안한 방법의 고주파대역의 LPC 분석

그림 6의 (a), (b)는 QMF 필터를 통과한 후에 저주파 및 고주파대역에서의 각각 프리-엠퍼시스를 통해 보상된 신호의 LPC로 비교 분석한 것이다. 그림 5와 6을 분석한 결과 제안한 방법에 있어서 약화된 고주파 신호가 강조됨을 알 수가 있다.

그림 7과 8은 유성음 프레임의 경우와 마찬가지로, 제안한 방법에 의하여 무성음에 해당하는 프레임의 고주파 성분을 보상한 후 그래프로 나타난 그림이다. 그림 7과 8은 성도모델을 잘 표현할 수 있는 LPC로 비교 분석한 것이다. 그림 7의 (b),(c)를 비교 분석한 결과 유성음 프레임과 마찬가지로 제안한 방법을 사용한 경우 고주파 영역이 보상됨을 알 수가 있다. 즉, 제안한 알고리즘은 고주파 영역의 약화된 음성신호를 보상하여 평탄화 함으로써, 효과적으로 성도특성을 추출할 수 있다. 그러므로 제안한 평탄화 알고리즘은 음성인식, 음성 합성 및 분석 등의 음성신호처리 분야에 광범위하게 적용되어질 수 있다.

6. 결론

음성신호를 관찰하였을 때 성문특성으로 인해서 고주파 쪽 특성이 약화되는 경향이 있다. 약화된 고주파 특성을 보상하기 위하여 프리-엠퍼시스 필터를 통해 보상하고 있다. 프리-엠퍼시스 필터를 간단한 수식으로 표현하면 $y(n) = s(n) - As(n-1)$ 와 같이 차분 방정식으로 나타낼 수 있다[3,4] 여기서 A값은 보통 0.9에서 1사이의 값을 사용하고 있으며 기존방법으로는 A값을 고정시켜서 보상하는 방법과 $A = R_n(1) / R_n(0)$ 을 대입하여 성도특성을 보상하는 방법이 있다. 그러나 음성의 스펙트럼에서 전체 주파수 중에 양 끝 부분의 주파수영역은 잘 보상되나 중간영역에 위치한 주파수 부분은 정확히 보상되지 않는다.

QMF 필터를 이용하여 입력신호를 고주파와 저주파의 2개의 대역으로 분할하여 각 밴드에 프리-엠퍼시스 필터를 적용하여 약화되어진 특성을 정확히 보상할 수 있다. 본 논문에서는 기울기 A값 조절을 통한 프리-엠퍼시스 필터 성질을 고려하여 전처리에 QMF 필터를 사용하여 약화되어진 특성을 정확히 보상하는 새로운 방법을 제안한다. 주파수 영역의 포락선 기울기를 평탄화 해줌으로서 SNR을 높이고 음질향상 및 인식에 유효하다. 이 방법은 일반적으로 A/D 변환을 위한 저역 통과 필터링에 앞서서 20dB/decade 정도의 고주파 영역을 강조하는 역할을 하며 성도 추출을 정확히 할 수 있다. 실험 결과, 각각의 프레임 구간 특성에 따라 고주파 특성, 혹은 저주파 특성을 강조함을 알 수가 있는데 분석을 위해 LPC 분석을 수행한 결과, 성도특성을 제대로 tracking하고 있음을 알 수 있다. 그러므로 본 논문에서 제안한 평탄화 알고리즘은 음성인식, 음성 분석 및 합성 등, 음성신호처리 분야에 광범위하게 적용되어질 수 있다.

참고문헌

- [1] L. R. Rabiner, R.W. Schafer, *Digital Processing of Speech Signal*, Prentice Hall, 1978.
- [2] 오영환, *음성언어정보처리*, 홍릉과학 출판사, 1998.
- [3] Douglas O. shaughnessy, *Speech Communication*, IEEE Press, 1996.
- [4] L. R. Rabiner & Biing-Hwang Juang, *Fundamentals Of Speech Recognition*, Prentice-Hall AT&T, U.S.A., 1993.
- [5] Thomas Parsons, *Voice and Speech Processing*, McGraw Hill.

- [6] 민소연, 강은영, 배명진, “성문특성이 제거된 성도특성에 관한 연구”, 대한전자공학회, 추계 종합학술대회, 2000년 11월 25일.
- [7] Oppenheim, Schafer, *Discrete Time Signal Processing*, Prentice Hall, 1989. AT&T, U.S.A, 1993.
- [8] 강은영, 민소연, 배명진, “개선된 피치 검출을 위한 스펙트럼 평탄화 기법에 관한 연구”, 한국음향학회지, 제21권 제3호, 2002년 4월.
- [9] 민소연, 장경아, 배명진, “선형스펙트럼쌍을 이용한 성문특성이 제거된 성도특성 추출법에 관한 연구”, 한국음향학회지, 제21권 제7호, 2002년 10월.

민 소 연(So-Yeon Min)

[종신회원]



- 1993년 2월 : 송실대학교 전자공학과 (공학사)
- 1995년 2월 : 송실대학교 전자공학과 (공학석사)
- 2003년 2월 : 송실대학교 전자공학과 (공학박사)
- 2005년 3월 ~ 현재 : 서일대학 정보통신과 전임강사

<관심분야>

음성코딩, 음성신호처리, 멀티미디어 통신