

선형 스펙트럼쌍을 이용한 성문특성이 제거된 성도특성 추출법에 관한 연구

A Study on Extraction of Vocal Tract Characteristic After Canceling the Vocal Cord Property Using the Line Spectrum Pairs

민 소 연*, 장 경 아**, 배 명 진**
(So-Yeon Min*, Kyung-A Jang**, Myung-Jin Bae**)

* 송실대학교 전자공학과, ** 송실대학교 정보통신전자공학부
(접수일자: 2002년 8월 29일; 채택일자: 2002년 10월 1일)

프리엠퍼시스 필터의 일반적인 형태는 $y(n) = s(n) - As(n-1)$ 이고, 여기서 A값은 유성음의 경우 0.9~1.0 사이의 값이다. 또한 A값은 프리엠퍼시스의 기울기 값을 반영하고 기존의 방법에서는 자기상관계수 값인 $R(1)/R(0)$ 를 사용한다. 본 논문에서는 성문특성으로 인해 고주파 특성이 약화되는 것을 보상하기 위하여 새로운 평탄화 기법을 제안한다. 우선 포먼트 주파수 예측을 위해 LSP 파라미터의 간격정보를 사용하였다. 찾아진 포먼트 주파수들간의 선형보간을 통해 기울기와 역기울기 값을 구하여 평탄화 과정을 수행한다. 실험결과에서는 제안한 방법이 기존의 방법보다 평탄화 특성이 우수한 것으로 나타났다. 즉 본 논문에서는 약화된 고주파 성분을 보상하는 과정에서 평탄화 요소로 LSP의 간격정보를 사용하였다.

핵심용어: 선형스펙트럼쌍, 성문특성, 성도특성, 평탄화, 자기상관법

주요분야: 음성처리 분야 (2, 4)

The most common form of pre-emphasis is $y(n) = s(n) - As(n-1)$, where A typically lies between 0.9 and 1.0 in voiced signal. Also, 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 interval information of LSP to estimate formant frequency. After obtaining the value of slope and inverse slope using linear interpolation among formant frequency, flattening process is followed. Experimental results show that the proposed method flattened the weakened high frequency components effectively. That is, we could improve the flattening characteristics by using interval information of LSP as flattening factor at the process that compensates weakened high frequency components.

Keywords: Line spectrum pairs, Vocal cord characteristics, Vocal tract characteristics, Flattening, Auto-correlation method

ASK subject classification: Speech signal processing (2, 4)

I. 서론

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 음성정보를 추출하거나 강조할 수가 있다. 따라서 음성

신호의 생성에 대한 수학적 모델은 음성을 처리하는데 있어서 매우 중요하다. 본 논문에서는 주파수 영역에서 LSP를 이용하여 포먼트 주파수를 추출하고 이 과정을 통해 얻은 포락선 기울기를 사용하여 이에 따라 A값을 각 프레임별로 특성에 맞게 보상해 줄 것을 제안한다. 2장에서는 음성생성 모델에 대하여 설명하고 3장에서는 기존의 자기 상관법을 이용한 기울기 측정법과 본 논문에서 제안한 주파수 영역의 포락선 기울기 측정법을 수식을 통해서 살펴본다. 4장, 실험 및 결과에서는 기존의 방법과 제안한 방법에 있어서 측정된 기울기와 평탄화된 특성에 대하여 설명하고 5장에서는 결론을 맺는다.

II. 성문특성을 고려한 음성생성모델

그림 1은 성문특성을 고려한 음성생성 모델이다. 그림 1에서 $E(z)$ 은 여기신호, $G(z)$ 은 2 pole로 모델링이 가능한 성문특성, $V(z)$ 은 성도특성, 즉 주파수 영역에서의 포먼트 특성을 나타낸다. 그리고 $L(z)$ 은 1 zero로 모델링이 가능한 입술의 방사특성을 나타내는데 마이크로폰을 가깝게 대고 발음을 한 경우에 있어서는 커플링(Coupling)이 발생하여 zero 특성이 없어지게 된다. 마이크로폰을 멀리 대고 발음한 순간에는 그림 1에서 성문특성을 나타내는 2 pole 중 하나가 1 zero의 $L(z)$ 와 상쇄되어 없어지므로 성문특성에 의해 영향을 받는 성도특성을 제대로 관찰하기 위해서는 그림 1에서처럼 $L_1(z)$ 을 사용하여 보상을 시켜주어야 한다. 위에서 설명한 내용을 간단히 수식을 통해 살펴보면 다음과 같다.

$$S(z) = E(z) \cdot G(z) \cdot V(z) \cdot L(z) \tag{1}$$

식 (1)은 마이크로폰을 멀리 대고 발성을 한 경우를 나타낸다. 그러므로 2 pole로 모델링한 성문특성이 1 zero로 모델링되는 입술의 방사특성으로 인해 전체적으로 1 pole로 나타낼 수가 있고, 대략적으로 주파수 영역에서의 기울기는 -20 dB/decade 정도로 나타난다[1-3].

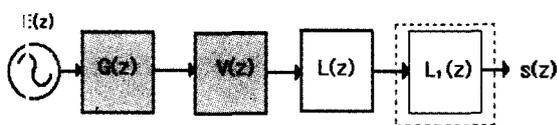


그림 1. 본 논문에서 고려한 음성생성모델
Fig. 1. Considered model for speech production in this paper.

$$S'(z) = L_1(z) \cdot S(z) = E(z) \cdot V(z) \tag{2}$$

식 (2)는 성도특성의 정확한 분석을 위해 성도특성에 미치는 성문특성을 제거하기 위한 경우를 나타낸 식이다. 식 (2)는 다음과 같이 표현 가능하다.

$$V(z) \cong \frac{S'(z)}{E(z)} \tag{3}$$

즉, 식 (1)~식 (3)을 통해서 마이크로폰의 위치에 따라 주파수 영역의 포먼트 특성의 기울기가 변화함을 알 수 있다.

III. 제안한 주파수영역에서의 기울기 측정법

3.1. 자기상관법을 이용한 기울기 측정

단구간 자기상관 함수는 식 (4)로 표현 가능하다[4-6].

$$\begin{aligned} \phi_n(i, j) &= \sum_{m=0}^{N-1-(i-j)} s_n(m) s_n(m+i-j), \\ 1 \leq i \leq p, 0 \leq j \leq p \end{aligned} \tag{4}$$

여기서,

$$R_n(j) = \sum_{m=0}^{N-1-j} s_n(m) s_n(m+j) \tag{5}$$

$$\sum_{j=1}^p a_j \phi_n(i, j) = \phi_n(i, 0), \text{ for } i=1, \dots, p \tag{6}$$

자기상관법을 이용하여 식 (6)을 풀면 다음과 같이 표현된다.

$$\begin{bmatrix} R_n(0) & R_n(1) & \dots & R_n(p-1) \\ R_n(1) & \cdot & \cdot & R_n(p-2) \\ \vdots & \vdots & \vdots & \vdots \\ R_n(p-1) & \cdot & \cdot & R_n(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ \vdots \\ R_n(p) \end{bmatrix} \tag{7}$$

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

$$a_1 = \frac{R_n(1)}{R_n(0)} \tag{8}$$

3.2. 제안한 방법의 기울기 추정

그림 2는 본 논문에서 기울기 측정을 위해 제안한 방법의 블록도이다. LSP를 구하는 과정을 보면 음성신호를 입

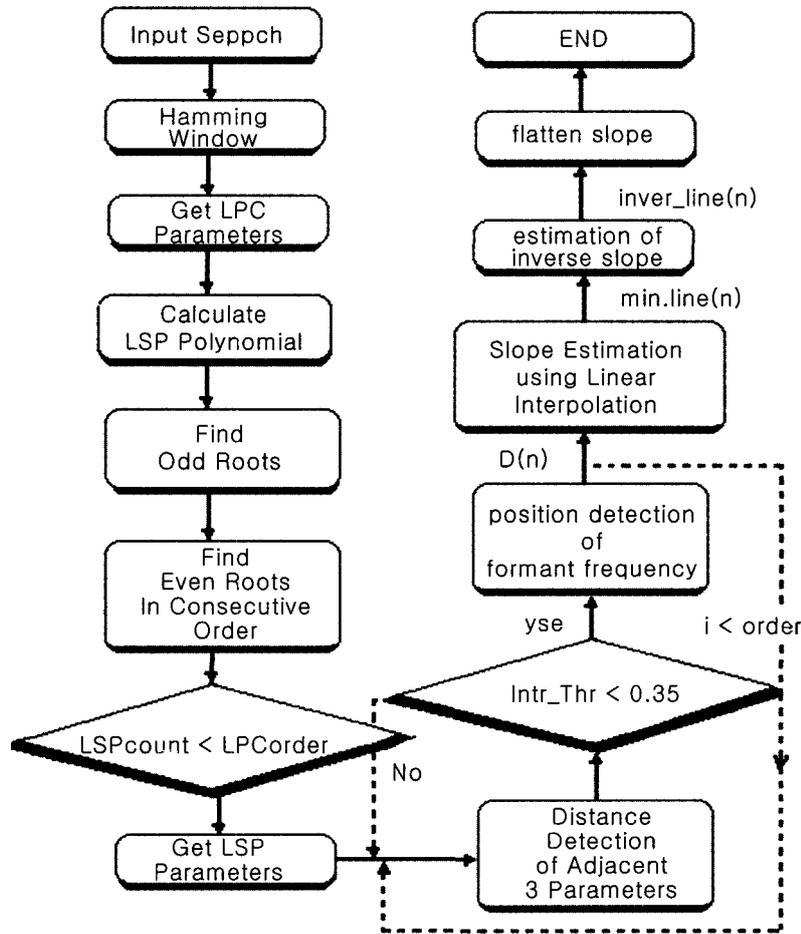


그림 2. 제안한방법의 블록도
Fig. 2. The blockdiagram of the proposed method.

력받고 각 프레임마다 해밍 윈도우를 취한다. 윈도우를 취한 데이터를 이용해 LPC (Linear Predictive Coding) 계수를 찾아내고, 이후에 LSP 파라미터를 얻어낼 수 있다. 10차의 LSP 파라미터는 각각 주파수 위치를 나타내고 있으며 두 파라미터의 위치간격으로 포먼트가 위치한 주파수 위치를 알아낼 수 있다. 즉, 제안한 방법에서는 3개의 LSP 파라미터를 하나의 단위로 하여 첫 번째와 두 번째의 간격정보, 두 번째와 세 번째의 간격정보의 값을 추출한다. 이것은 아래 식 (9-b), (9-c)에 해당한다. 또한 식 (9-a)는 첫 번째 파라미터와 세 번째 파라미터의 간격 정보를 나타낸다.

$$LSP(i) = LSP(i+2) - LSP(i) \quad (9-a)$$

$$LSP_2(i) = LSP(i+2) - LSP(i+1) \quad (9-b)$$

$$LSP_1(i) = LSP(i+1) - LSP(i), 1 < i < 8 \quad (9-c)$$

$$D_2(i) = \frac{LSP_2(i)}{LSP(i)} \quad (10-a)$$

$$D_1(i) = \frac{LSP_1(i)}{LSP(i)}, 1 < i < 8 \quad (10-b)$$

계산에서 얻어진 D_1 과 D_2 중에서 문턱값보다 작은 값을 선택하고, 그 주파수 간격 정보를 알아내어 저장한다. 이를 순차적으로 반복하여 포먼트의 위치정보를 얻어낼 수 있다. 문턱 값보다 작은 밴드폭을 가진 인접한 두 LSP 파라미터를 $LSP_{(def)}$ 라 하면 식 (11)은 포먼트 주파수에 해당한다.

$$D(n) = \frac{LSP_{(def)}}{2} \quad (11)$$

위의 과정을 통해 얻어진 포먼트 주파수의 개수가 n 개라고 가정하고, 주파수 영역에서의 좌표를 나타내면

(D(1), DB(D(1)), (D(2), DB(D(2))), (D(3), DB(D(3))), ..., (D(n), DB(D(n)))이다.

$$| (m) = DB(D(n)) + (m - D(n))f'(D(n)) + \frac{(m - D(n))^2}{2} f''(D(n)) \dots \quad (12-a)$$

$$f'(D(n)) = \frac{DB(D(n+1)) - DB(D(n))}{D(n+1) - D(n)} \quad (12-b)$$

식 (12-a), (12-b)를 이용한 선형보간 과정을 거쳐 주파수영역에서 정확한 선형적 기울기를 추정할 수 있다 [7]. 구해진 기울기를 이용하여 스펙트럼 기울기를 평탄화 시키기 위해서는 역 기울기인 식 (13-b)를 이용한다. 또한, 식 (13-c)는 주파수 영역에서 평탄화된 스펙트럼의 결과를 나타낸다.

$$\min data = \min(DB(D(n))) \quad (13-a,b,c)$$

$$inver - | (m) = | (m) - 2(| (m) - \min data)$$

$$flatten(m) = DB(m) + inver - | (m), 0 < m < frame$$

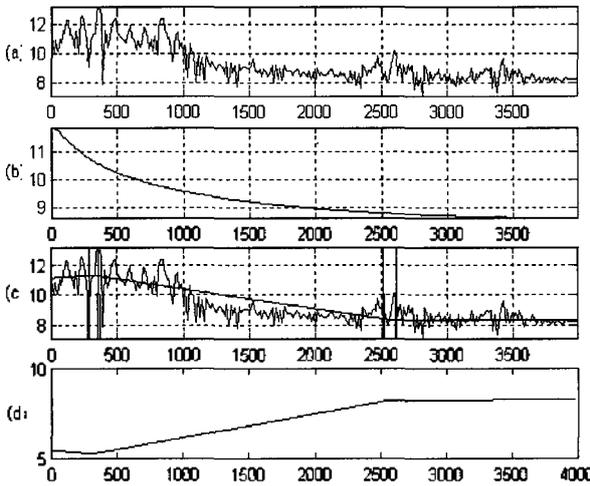


그림 3. 유성음의 기울기 측정 결과
 (a) 음성신호의 스펙트럼분석
 (b) 자기 상관법을 이용한 기울기측정
 (c) 제안한 방법을 이용한 기울기 측정
 (d) 제안한 방법의 역기울기 측정
 Fig. 3. Measured slope in voiced signal.
 (a) Spectrum analysis of speech signal,
 (b) measured slope by auto-correlation method,
 (c) measured slope by proposed method,
 (d) measured inverse slope by proposed method.

IV. 실험 및 결과

본 논문에서 제안한 방법의 실험을 위해 일기예보 뉴스에서 발췌한 남/녀 아나운서 음성을 8 kHz로 표본화하고 16비트로 양자화하여 사용하였다. 한 프레임의 길이는 256샘플이며 계수는 10차를 써서 특징벡터를 추출하였다. 그림 3에서 (a)는 주어진 음성시료 중 유성음에 대한 주파수 영역의 특성을 나타내고, (b)는 기존의 자기상관법을 이용하여 측정된 기울기이다. (c)는 제안한 방법의 실험결과로서 LSP를 추출하여 포만트위치를 추정하고 선형보간에 의해 기울기 측정법을 이용하여 구한 기울기와 주파수 영역의 특성을 나타내고, (d)는 스펙트럼 평탄화를 위한 역 기울기를 구한 그래프이다. 그림 4는 하모닉스를 제거하고 스펙트럼 포락선의 변화결과를 알기 위해 LPC분석을 수행한 결과이다. 그림 4의 (c), (d)를 비교한 결과를 보면 제안한 방법에 있어서 평탄화가 잘됨을 알 수가 있다. 그림 5는 기존의 방법과 제안한 방법간의 평탄화된 스펙트럼을 비교 분석한 결과이다. 그림 6은 무성음에 해당하는 프레임을 제안한 방법에 의해 측정된 기울기를 나타낸 그림이고, 그림 7은 기존의 방법과 제안한 방법간의 평탄화된 스펙트럼 포락선 변화를 비교하기 위하여

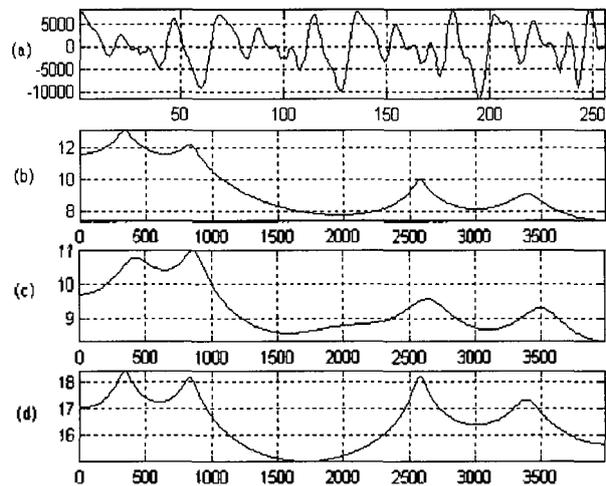


그림 4. 유성음의 LPC 분석 결과
 (a) 음성신호
 (b) 원음성의 LPC분석
 (c) 자기 상관법을 이용한 평탄화
 (d) 제안한 방법을 이용한 평탄화
 Fig. 4. LPC analysis of voiced signal.
 (a) Speech signal,
 (b) LPC analysis of speech signal,
 (c) flattened result by auto-correlation method,
 (d) flattened result by proposed method.

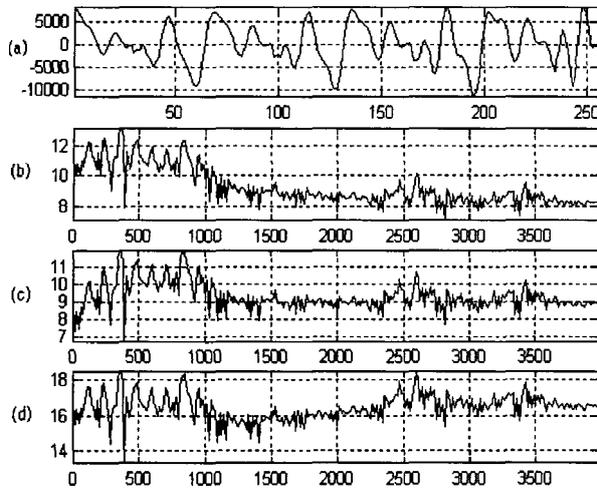


그림 5. 유성음의 스펙트럼 분석 결과

- (a) 음성신호
- (b) 원음성의 스펙트럼
- (c) 자기 상관법을 이용한 스펙트럼
- (d) 제안한 방법을 이용한 스펙트럼

Fig. 5. Spectrum analysis of voiced signal.

- (a) Speech signal,
- (b) spectrum analysis of speech signal,
- (c) spectrum analysis by auto-correlation method,
- (d) spectrum analysis by proposed method.

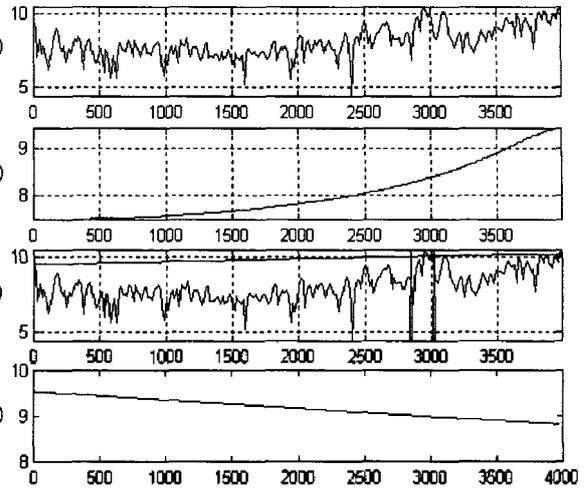


그림 6. 무성음의 기울기 측정 결과

- (a) 음성신호의 스펙트럼분석
- (b) 자기 상관법을 이용한 기울기 측정
- (c) 제안한 방법을 이용한 기울기 측정
- (d) 제안한 방법의 역기울기 측정

Fig. 6. Measured slope in unvoiced signal.

- (a) Spectrum analysis of speech signal,
- (b) measured slope by auto-correlation method,
- (c) measured slope by proposed method,
- (d) measured inverse slope by proposed method.

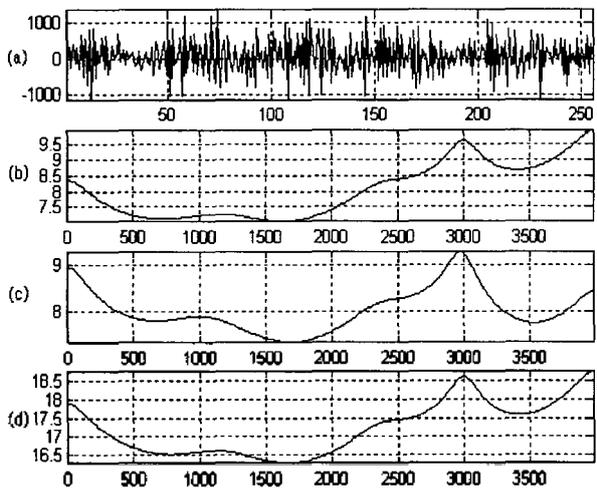


그림 7. 무성음의 LPC 분석 결과

- (a) 음성신호
- (b) 원음성의 LPC분석
- (c) 자기상관법을 이용한 평탄화
- (d) 제안한 방법을 이용한 평탄화

Fig. 7. LPC analysis of unvoiced signal.

- (a) Speech signal,
- (b) spectrum analysis of speech signal,
- (c) flattened result by auto-correlation method,
- (d) flattened result by proposed method.

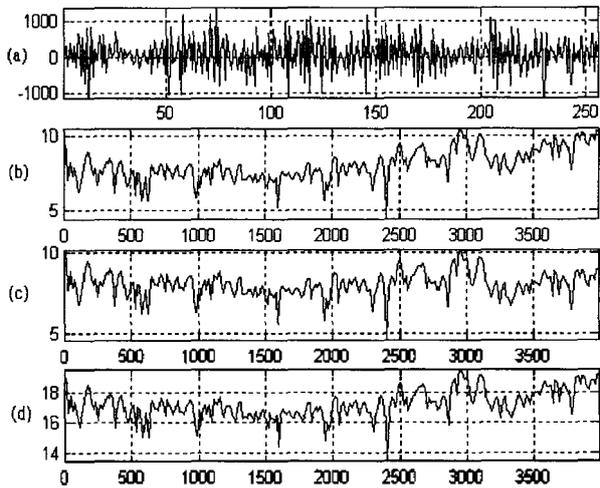


그림 8. 무성음의 스펙트럼 분석 결과

- (a) 음성신호
- (b) 원음성의 스펙트럼
- (c) 자기상관법을 이용한 스펙트럼
- (d) 제안한 방법을 이용한 스펙트럼

Fig. 8. Spectrum analysis of unvoiced signal.

- (a) Speech signal,
- (b) spectrum analysis of speech signal,
- (c) spectrum analysis by auto-correlation method,
- (d) spectrum analysis by proposed method.

LPC 분석을 수행한 결과이다. 그림 8은 제안한 방법에 의해 주파수 영역에서 보상된 결과를 나타내고 있다.

V. 결론

본 논문에서는 주파수 영역에서 LSP를 이용하여 순차적으로 포먼트 주파수를 찾아내어 기존보다 세밀한 선형화된 기울기를 추정하고 이를 이용하여 약화된 특성을 보상하는 새로운 방법을 제안한다. 주파수 영역의 포락선 기울기를 평탄화해 줌으로써 신호대 잡음비를 높이고 음질향상에 유효하다. 임의의 음성 프레임 구간에 대한 실험결과에서 제안한 방법을 이용한 경우, 마이크의 위치에 따른 변화나 음성 프레임 구간 특성을 적응적으로 반영하여 성문특성이 제거된 평탄화된 성도특성을 추출할 수가 있다.

감사의 글

본 연구는 2001년도 송실대학교 교내연구비 지원으로 이루어졌습니다.

참고 문헌

1. 배명진, "디지털 음성분석," 동영출판사, 38-42, 108-118, 1998.
2. Thomas Parsons, *Voice and speech processing*, McGraw-Hill, 262-267, 1986.
3. L. R. Rabiner and B.-H. Juang, *Fundamentals of speech recognition*, Prentice-Hall AT&T, 112-113, 1993.
4. A. M. Kondoz, *Digital Speech*, John Wiley & Sons, 44-46, 79-95, 1994.
5. D. O' shaughnessy, "Speech communication," *IEEE Press*, 188-191, 1996.
6. L. R. Rabiner, and R. W. Schafer, *Digital Processing of Speech Signal*, Prentice Hall, 38-48, 1978.
7. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, Cambridge University Press, 2nd Edition, 108-110, 1992.

저자 약력

● 민 소 연 (So-Yeon Min)

한국음향학회지 제21권 제3호 참조
현재: 송실대학교 전자공학과 박사수로

● 장 경 아 (Kyung-A Jang)

한국음향학회지 제21권 제2호 참조
현재: 송실대학교 정보통신공학과 박사 과정

● 배 명 진 (Myung-Jin Bae)

한국음향학회지 제21권 제3호 참조
현재: 송실대학교 정보통신 전자공학부 교수