

# 음성신호 전처리를 위한 에너지 의존 프리엠퍼시스

## Energy-Dependent Preemphasis for Speech Signal Preprocessing

김 동 준\*, 박 상 회\*\*  
(Dong-Jun Kim\*, Sang-Hui Park\*\*)

### 요 약

본 연구에서는 성문파에 의한 음원 특성과 입술에서 음성이 방사될 때 발생하는 방사 특성을 효과적으로 제거하기 위하여 성문파의 음원 특성과 입술에서의 방사 특성은 근사적으로 음파의 에너지와 비례한다고 가정하고, 정규화된 단구간 에너지를 이용하는 에너지 의존 프리엠퍼시스 기법을 제안하며, 이를 이용하여 비안정 구간인 발음의 시작 부분과 천이 구간에 대하여 음성 신호 분석 성능을 개선하고자 한다.

제안된 프리엠퍼시스 기법을 이용하여 5개 한국어 단모음의 스펙트럼 및 형성음 주파수 추출 등의 음성 신호 분석을 수행하고, 기존에 널리 이용되던 두 가지 프리엠퍼시스 기법과 성능을 비교하여 본 결과, 제안된 방법에 의한 스펙트럼의 형태가 기존의 방법에 비하여 상당히 개선되었고, 보다 더 정확한 형성음 주파수를 나타내며, 인접한 두 형성음 주파수가 중첩되는 현상이 제거되었음을 알 수 있었다.

### ABSTRACT

This study describes a modified preemphasis formula, what we call energy-dependent preemphasis(EDP). This uses the normalized short-term energy of speech signal, with the assumption that the source characteristics of the glottal pulses and the radiation characteristics of the lips are approximately proportional to the energy of speech signal.

Using this method, speech analyses, such as AR spectrum estimation, formant detection, are performed for nonstationary starting parts of 5 Korean single vowels. The results are compared with the conventional two preemphasis methods.

We found that the proposed preemphasis gave enhanced spectral shapes and more accurate formant frequencies and avoided overlapping phenomenon of adjacent two formants.

### I. 서 론

음성의 부호화(coding), 통신, 인식 및 합성 등 음성 신호처리의 전 영역에서 필수적인 전처리 과정인 프리엠퍼시스는 음성 신호의 정확한 모델링과 분석에 못지 않게 매우 중요하다. 대부분의 연구에서는 기존에 개발된 방법을 그대로 이용하고 있다. 일반적으로 가장 널리 이용되는 프리엠퍼시스 방법은 프리엠퍼시스 상수를 1 근처의 고정된 값으로 설정하거나, 데이터에 따라 자기 상관 함수의 비  $R(1)/R(0)$ 를 구하여 프리엠퍼시스하는 방법이다 [1, 2]. 그러나 이와 같은 방법은 발음이 정상 상태가 아닌 부분에서는 부정확한 결과를 산출할 가능성이 있다. Nakajima[2]는 적응 역필터(adaptive inverse filter)를 도입하여 스펙트럼 포락선을 편평하게 하는 연구를 발표하였

다. 이 접근 방식은 경험적인 것이고, 모음의 넓은 영역에 대하여 합당한 결과를 내었으나 방법이 매우 복잡하다.

효과적인 음성 신호의 분석을 위해서는 음성 발생 메카니즘의 생리학적인 연구로부터 추출된 기본 지식을 이용하는 것이 중요하고 또한 필연적이다. 음성은 인간의 발생 기관에서 발생된다. 현대의 음성 발생 이론에서는 음성 발생 과정을 음원의 발생, 조음, 그리고 음성의 방사로 구성된다고 가정한다. 음원은 음파의 에너지원이며 혀로부터 공급된 공기 흐름에 의하여 발생되며, 유성음은 성문의 주기적인 개폐 운동에 의하여 이루어지고, 무성음은 조음체에 의한 공기 흐름의 방해나 차단에 의하여 형성된다[3].

본 연구에서는 음성 발음시 발음의 시작 부분과 정상 상태가 되기 전인 천이 구간에 대하여 성문파에 의한 음원 특성과 입술에서의 방사 특성을 효과적으로 제거함으로써 보다 정확한 분석을 가능하게 하기 위하여 음원 특성과 방사 특성은 근사적으로 음파의 에너지와 비례한다고 가정하고, 정규화된 단구간 에너지(normalized short-term

\* 청주대학교 정보통신공학과

\*\* 연세대학교 전기공학과

접수일자: 1996년 8월 23일

energy)를 이용한 에너지 의존 프리엠퍼시스(energy-dependent preemphasis, EDP) 기법을 제안하며, 이를 이용하여 음성 신호 분석 성능을 개선하고자 한다.

## II. 음성 신호 분석

### 2.1 선형 예측 분석

음성의 음향적 특성을 가장 성공적으로 나타낸 모델 중의 하나는 Fant에 의해 개발된 선형 음성 발생 모델(linear speech production model)이다[1]. 음성 발생 모델의 매개변수들은 선형 예측 분석을 통하여 쉽게 얻어질 수 있다는 중요한 특성 때문에 선형 예측 모델은 음성 발생 모델과 잘 관련지어진다. 선형 음성 발생 모델은 그림 1과 같다.

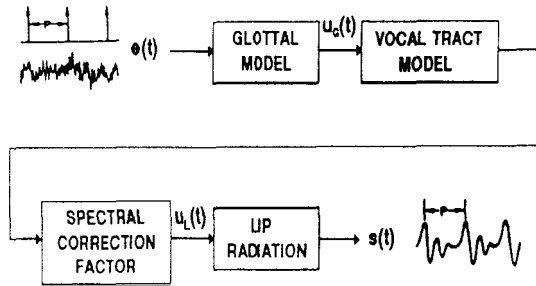


그림 1. 선형 음성 발생 모델[1]  
Fig. 1. A linear speech production model

이 모델의 분석 모델은 식 (1)과 같다.

$$E(z) = S(z) A(z) \quad (1)$$

여기서,  $E(z)$ 는 여기 모델,  $S(z)$ 는 음성 신호의 z-변환이고,  $A(z)$ 는 역필터로서, 근사적으로 성문과 형성 모델, 성도 모델, 입술의 방사 모델 등 세 모델의 곱의 역이다. 식 (1)을 표본화된 데이터 영역에서 표현하면

$$\begin{aligned} e(n) &= \sum_{i=0}^M a_i s(n-i) \\ &= s(n) + \sum_{i=1}^M a_i s(n-i) \\ &= s(n) - \hat{s}(n-1) \end{aligned} \quad (2)$$

이고, 여기서  $-a_i$ 는 예측기 계수(predictor coefficients),  $\hat{s}(n)$ 은 예측된 샘플이다.

본 연구에서는 선형 예측 분석을 위하여 적응 필터의 일종인 RLSL 알고리즘을 이용한다.

적용 필터는 그 동작을 회귀 알고리즘(recursive algorithm)에 의존하여 수행하며, 신호가 안정한 경우에는 알고리즘의 연속적인 반복 수행후에 최적해에 수렴한다. 비안정 신호의 경우 회귀 알고리즘은 입력 데이터의 통

계적 성질이 시간에 따라 변하는 것을 추적하는 추적 능력(tracking capability)을 나타낸다.

RLSL 알고리즘의 중요한 장점은 구조가 모듈(module)로 되어 있다는 것이다. 따라서 예측 차수를 증가시킬 때 모든 이전 값을 다시 계산할 필요가 없다. 또한 하드웨어 구현시 VLSI 기술에 적용이 용이하다. 그리고 lattice 구조는 음성발생모델과 일치되는 특성을 갖고 있다[4].

RLS lattice 필터들의 다양한 종류들 중에서 반사계수들을 직접적으로 update하는 오차 피이드백 구조는 장기간 수치적 안정성과 정확성을 갖는다. 그러나 이러한 구조는 매 시간마다  $22M$ 의 연산을 갖으며, 이는 기존의 a posteriori RLS lattice가  $16M$ 의 연산을 갖는 것과 비교할 때 높은 계산적 복잡성을 갖는다. 이러한 높은 계산적 복잡성을 기존의 a posteriori lattice와 같은  $16M$ 의 연산으로 줄이고, 또한 수치적 안정성과 정확성을 유지하기 위해 반사계수의 직접적인 update를 이용하는 새로운 적용 RLSL 알고리즘인 double/direct RLSL 알고리즘이 Orfanidis에 의해 제안되었으며, 본 연구에서는 이를 이용한다[5].

### 2.2 적응 AR 스펙트럼 분석

스펙트럼 추정에서 실용적으로 잘 이용되는 모델은 AR (autoregressive) 모델이며, AR 모델이 시변이면 모델의 매개변수들도 시간에 의존적이 되어  $a_1(n), a_2(n), \dots, a_M(n)$ 으로 표기된다. 이 경우 시변 AR 프로세스의 전력 스펙트럼은 다음과 같다.

$$S_{AR}(\omega, n) = \frac{\sigma_v^2}{|1 + \sum_{i=1}^M a_i(n)e^{-j\omega i}|^2} \quad (3)$$

본 연구에서 이용한 double/direct RLSL 알고리즘은 lattice 구조에서 개발된 알고리즘이므로 반사계수가 직접 추출되므로 AR 매개변수를 추출하기 위해서는 다음과 같은 forward Levinson recursion을 이용한다.

$$A_{p+1}(z) = A_p(z) - \gamma_{p+1} z^{-(p+1)} A_p(z^{-1}) \quad (4)$$

대수 스펙트럼(log spectrum)은 식 (5)로 계산된다.

$$L_{AR}(\omega, n) = 20 \text{Log}_{10}(S_{AR}(\omega, n)) \quad (5)$$

구해진 대수 스펙트럼으로부터 첨두치를 찾아내어 형성 음 주파수를 구한다. 그러나 이러한 첨두치 추출 방법에서는 대수 스펙트럼이 뚜렷한 첨두치가 아닌 변곡점으로 나타나는 경우, 일반적인 첨두치 추출 방법에 의해 형성음을 구하면 그 일부가 유실될 수 있다. 따라서 본 연구에서는 1, 2차 미분을 이용한 첨두치 추출방법을 사용하여 형성음 주파수를 찾아낸다[6].

### Ⅲ. 에너지 의존 프리엠퍼시스

프리엠퍼시스는 성도의 공진 주파수와 대역폭에 영향을 미치는 다른 인자들을 제거해 주므로 스펙트럼상에서 낮은 진폭과 높은 주파수를 갖는 형성음 주파수물 모델링하는데 좋은 결과를 낼 수 있다. 이는 프리엠퍼시스에서 음원의 특성과 방사 특성을 제거함으로써 성도 자체의 특성이 잘 반영되도록 처리해 주기 때문이다. 또한 프리엠퍼시스는 유성음에서 음원에 의한 스펙트럼의 fall-off 현상에 대처하기 위해 zero를 첨가하므로 스펙트럼의 dynamic range를 줄여 주고 고정점 연산을 쉽게 해준다. 프리엠퍼시스에서  $\mu$ 가 큰 값이면 zero가 매우 낮은 주파수에 위치하므로 제 1 형성음 주파수 F1 이하의 영역에서 상당한 감쇄가 유발되고,  $\mu$ 값이 작으면 F1 이하의 영역에서 상대적으로 작은 감쇄가 있게 된다.

프리엠퍼시스에서 이용되는 필터는 1차 미분 역 필터(first-order differentiation/difference inverse filter), 적응 1차 역 필터(adaptive first-order inverse filter), 적응 다단계 1차 역 필터(adaptive multistage first-order inverse filter) 등이 있다[4]. 1차 미분 역 필터는 단순히 A/D 변환된 신호의 미분을 구해주는 것이며, 적응 1차 역 필터는 음성 신호를 다음과 같이 처리한다.

$$y(n) = s(n) - \mu s(n-1) \tag{6}$$

여기서,  $\mu$ 는 프리엠퍼시스의 정도를 제어하는 값으로 single zero 필터의 차단 주파수(cutoff frequency)를 정한다.  $\mu$ 는 0.94, 1 등으로 고정하는 경우가 많으며, 프리엠퍼시스 후 스펙트럼의 편평도(flatness)를 극대화하는 측면에서  $\mu = R(1)/R(0)$ 을 최적 선택으로 하기도 한다. 이와 같이  $\mu$ 를 선택하면, 대부분의 유성음에서는  $\mu \approx 1$ 이 되어 근사적인 미분기가 되고 무성음에서는 상대적으로 작은  $\mu$ 를 갖게 되어 프리엠퍼시스가 영향을 덜 미치게 된다. 또한 현재 처리되고 있는 테이타의 자기 상관 함수에 근거하여  $\mu$ 를 결정하므로 고정된  $\mu$ 의 경우보다는 발음의 시작 부분에 대한 반응이 잘 될 수 있다. 그러나 보통  $R(1)$ 과  $R(0)$ 의 값은 큰 차이를 나타내지 않으므로  $\mu$ 값은 처리되는 음성 테이타에 의존하면서 최대값 1근처에서 결정될 것이다.

성문파는 유성음이 발생하는 순간부터 작은 진폭으로부터 시작하여 정상상태에 이르는 것을 알 수 있다[3]. 이때 성문파의 크기는 성도를 흐르는 공기의 부피 속도에 영향을 미치므로 방사된 음파의 에너지와 근사적으로 비례한다고 볼 수 있다. 그림 2는 음성 발생시 각 단계마다의 주파수 특성을 종합하여 나타낸 것이다[1, 2].

유성음은 발음의 정상 상태에서 약  $-12[\text{dB}/\text{oct}]$ 의 음원 특성과  $0[\text{dB}/\text{oct}]$ 의 성도 특성, 그리고 약  $6[\text{dB}/\text{oct}]$ 의 방사 특성이 결합되어 입술로부터 방사된 음성 신호는 약  $-6[\text{dB}/\text{oct}]$ 의 스펙트럼 경사를 나타내게 된다. 무성음의 경우에는 음원의 특성이  $0[\text{dB}/\text{oct}]$ 이므로 성도 특성과 방

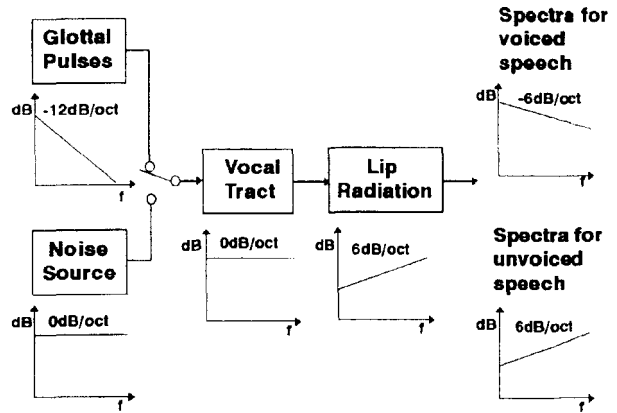


그림 2 음성 발생 과정의 특성  
Fig 2. Characteristics of speech production process

사 특성이 결합되면 출력되는 음성 신호의 스펙트럼 경사는 약  $6[\text{dB}/\text{oct}]$ 로 나타난다.

모음을 발음할 때 발음 시작 전에는 모든 특성이 나타나지 않는 상태이므로  $0[\text{dB}/\text{oct}]$ 라고 볼 수 있다. 성문파가 작은 진폭으로부터 발생하기 시작하면 서서히 음원 특성과 방사 특성이 나타나기 시작할 것이다. 성문파가 진폭이 작아서 공기 흐름의 양이 적을 때에는 음성 신호 자체의 진폭과 에너지도 작고, 음원 특성과 방사 특성도 정상 상태일 때의 스펙트럼 경사보다 작을 것이다. 따라서 발음의 시작 부분에서는 음성 신호의 스펙트럼 경사가  $0[\text{dB}/\text{oct}]$ 보다 크고  $-6[\text{dB}/\text{oct}]$ 보다는 작은 상태를 나타낼 것이다. 성문파가 정상 상태에 이르면 음원 특성과 방사 특성이 그림 2에서와 같이 나타나므로 스펙트럼이  $-6[\text{dB}/\text{oct}]$ 의 경사를 나타낼 것이다. 무성 자음의 발음시에는 순수 자음 부분에 대해서는 그림 2에 의하여  $6[\text{dB}/\text{oct}]$ 의 스펙트럼 경사 특성을 나타내지만, 성문파가 발생하기 시작하는 부분 이후에는 점차로 스펙트럼의 경사가  $0[\text{dB}/\text{oct}]$ 를 거쳐 부의 값으로 변화되어 후속 모음이 정상 상태일 때에는 결국  $-6[\text{dB}/\text{oct}]$ 가 될 것이다. 따라서 프리엠퍼시스의 정도를 성문파의 발생 시점부터 음성 신호의 진폭이나 에너지에 따라 비례적으로 변화시켜 주어야 음성 발생 과정의 모든 특성을 효과적으로 반영할 수 있을 것이다.

본 연구에서는 발음의 시작 부분에서 보다 정확한 분석을 위하여 음성 발음시 성문파에 의한 음원 특성과 입술에서의 방사 특성은 근사적으로 음파의 에너지와 비례한다고 가정하여, 정규화된 단구간 에너지를 이용한 에너지 의존 프리엠퍼시스 기법을 제안한다. 에너지 의존 프리엠퍼시스 필터는 다음과 같이 표현될 수 있다.

$$F_i(z) = (1 - \mu_i z^{-1}) \tag{7}$$

여기서,  $\mu_i$ 는  $i$ 번째 프레임(frame)의 역 필터 계수이며, 다음과 같이 정해진다.

단구간 에너지를 구하는 프레임의 크기를  $N$ 이라 하면 단구간 에너지는 식 (8)에 의하여 구할 수 있다.

$$E_n = \sum_{m=n-N+1}^n x^2(m) \quad (8)$$

여기서,  $x(m)$ 은  $m$ 번째 음성 데이터이고,  $n$ 은 전체 음성 데이터 수를 나타낸다.

식 (8)로부터  $\mu_i$ 는 다음과 같이 결정된다.

$$\mu_i = \frac{E_x}{E_{max}} \quad (9)$$

여기서,  $E_{max}$ 는 계산된 단구간 에너지 중에서 최대값을 의미한다.

### IV. 실험 및 결과 고찰

#### 4.1 실험

실험의 전 과정에 대한 블록 선도는 그림 3과 같다.

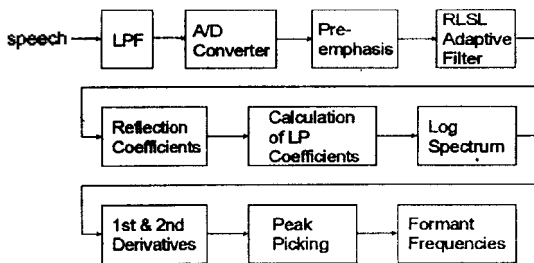


그림 3 음성 분석 과정의 블록 선도  
Fig 3. Block diagram of speech analysis

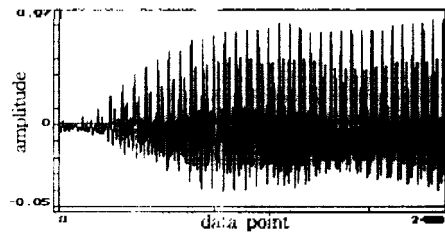
녹음된 음성 신호는  $f_c \cong 4.7$ [kHz]의 LPF를 통과하여 10[kHz]로 샘플링되고, 프리엠퍼시스된 후 10차로 구성된 RLSL 적응 알고리즘에 의하여 반사계수가 계산된다. 이 반사계수를 이용하여 음성신호의 분석이 이루어진다.

프리엠퍼시스 과정에서 본 연구에서 제안한 방법과 기존에 널리 이용되는 두 방법, 즉 전형적인  $\mu(=0.94)$ , 적응적  $\mu(=R(1)/R(0))$ 가 적용되어 각각의 성능이 비교된다.

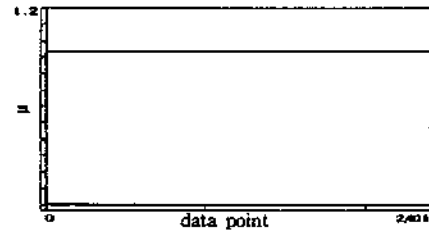
#### 4.2 결과 고찰

그림 4는 모음 /오/의 원 파형과 전형적인  $\mu$ , 적응적  $\mu$ , 그리고 본 연구에서 제안한  $\mu$ 의 궤적을 나타낸 것이다.

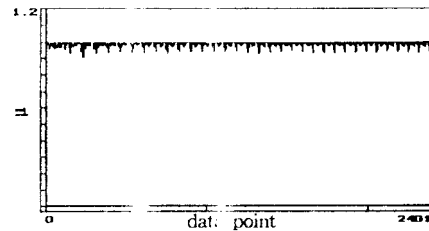
제안한 방법의  $\mu$ 는 발음 초기에 매우 작은 값으로 나타나다가 특정 문턱치에 이르면 정상 상태에 이른 것으로 보고  $\mu=1$ 로 고정하였다. 여기서 문턱치는 실험적으로 0.7로 정하였다. 그림 5는  $\mu=0.94$ 일 때와 본 연구에서 제안한 EDP일 때의 분석에서 한 화자의 모음 /아/에 대하여 발음의 시작 부분과 천이 구간의 스펙트럼을 나타낸 것이고, 그림 6은 나머지 4 모음의 시작 부분의 스펙트럼을 나타낸 것이다.



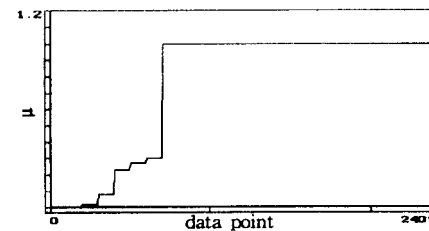
(a)/오/의 원파형



(b)  $\mu = 0.94$



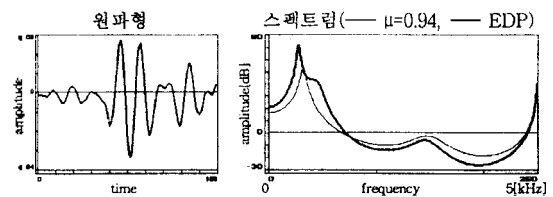
(c)  $\mu = R(1)/R(0)$



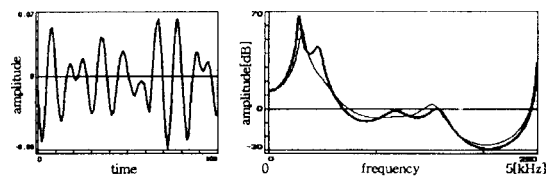
(d) 제안한 방법의  $\mu$

그림 4. /오/의 원파형과  $\mu$ 의 궤적

Fig 4. Waveform and  $\mu$  trajectories for /오/



(a)/아/의 시작 부분



(b)/아/의 천이 구간

그림 5.  $\mu = 0.94$ 일 때와 EDP에 의한 /아/의 스펙트럼

Fig 5. Estimated spectra using  $\mu = 0.94$  and EDP for /아/

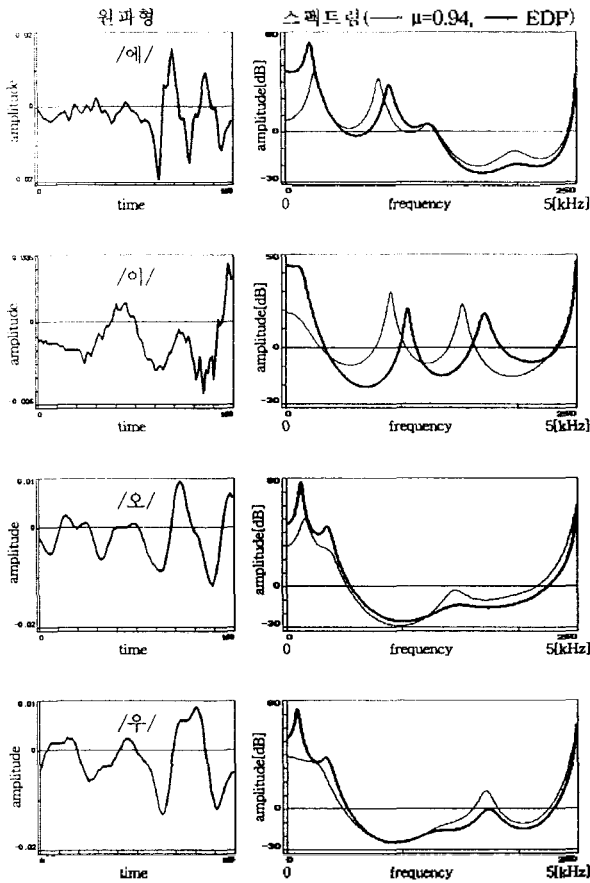


그림 6.  $\mu = 0.94$ 일 때와 EDP에 의한 4개 모음 시작부분의 스펙트럼  
 Fig 6. Estimated spectra using  $\mu = 0.94$  and EDP for 4 vowels in start

그림 7에서는  $\mu = R(1)/R(0)$ 일 때와 EDP일 때의 5개 단모음에 대한 시작 부분의 스펙트럼을 나타내었다.

모음 /아/, /오/, /우/의 경우 기존의 두 방법에 비해 F2가 잘 나타나는 것을 알 수 있다. 또한 /에/, /이/의 경우 새 방법의 형성음의 위치가 상당히 다르게 나타나는 현상을 관찰할 수 있다. 이를 검증하기 위하여 발음의 앞부분을 이용하여 FFT의 결과와 비교하였으며, 그 결과를 그림 8~9에 나타내었다.

FFT와의 비교 결과, 본 연구에서 제안한 프리엠퍼시스 기법이 발음의 시작 부분과 천이 구간에서 보다 정확한 형성음 위치를 검출할 수 있다는 것을 알 수 있다.

대수 스펙트럼의 1, 2차 미분을 이용한 형성음 주파수 검출의 실험 결과 형성음 검출의 오차의 수를 표 1에 나타내었다. 오차의 검출 기준은 일반적인 형성음 주파수 위치보다 1[kHz] 이상 벗어난 경우, 검출되지 않는 경우, 그리고 평균보다 갑자기 커지거나 작아지는 경우를 오차로 규정하였다. 이 표는 각 모음에 대하여 4(화자)×3(발음)×18(프레임) = 216[회]의 형성음 검출 시도에서 오차를 나타낸다.

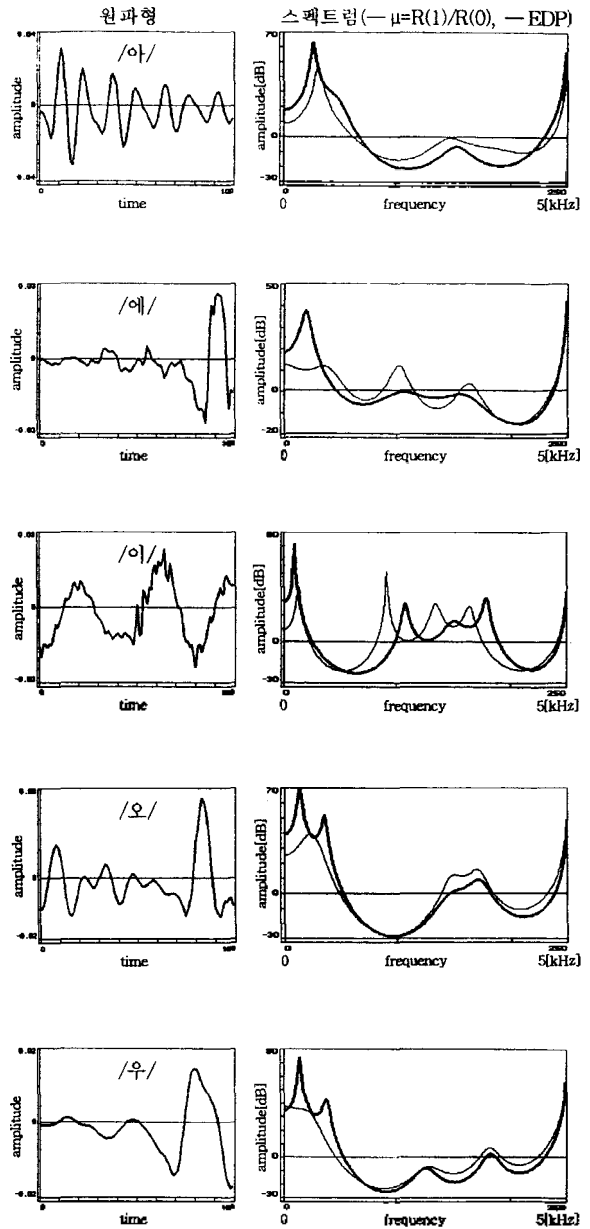
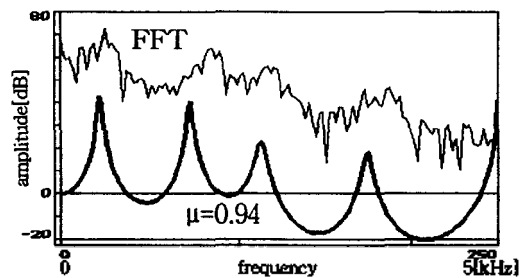
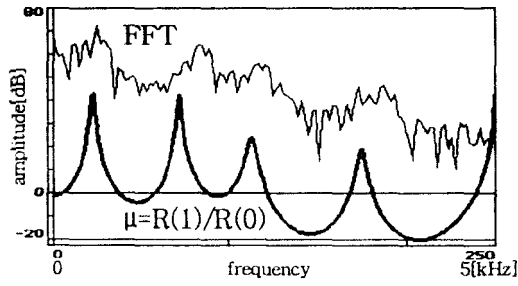


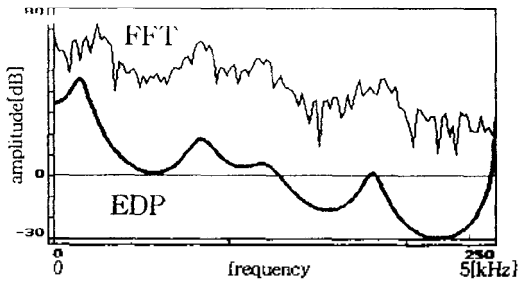
그림 7.  $\mu = R(1)/R(0)$ 일 때와 EDP에 의한 5개 모음 시작부분의 스펙트럼  
 Fig 7. Estimated spectra using  $\mu = R(1)/R(0)$  and EDP for 5 vowels in start



(a) FFT와  $\mu = 0.94$ 일 때의 스펙트럼



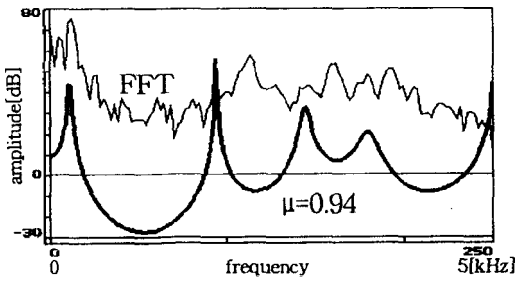
(b) FFT와  $\mu = R(1)/R(0)$ 일 때



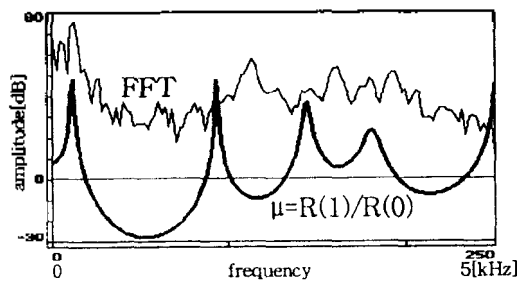
(c) FFT와 EDP일 때

그림 8. 세 가지 프리엠퍼시스에 대한 /에/의 시작 부분 스펙트럼 비교

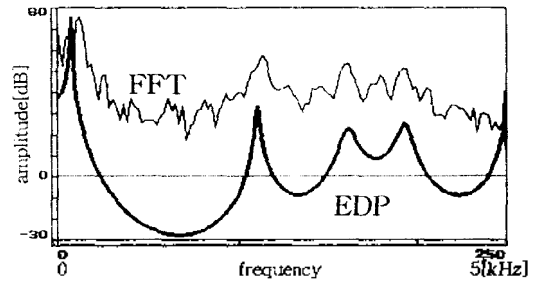
Fig 8. Spectra comparison of 3 preemphasis methods for start of /에/



(a) FFT와  $\mu = 0.94$ 일 때의 스펙트럼



(b) FFT와  $\mu = R(1)/R(0)$ 일 때



(c) FFT와 EDP일 때

그림 9. 세 가지 프리엠퍼시스에 대한 /이/의 시작 부분 스펙트럼 비교

Fig 9. Spectra comparison of 3 preemphasis methods for start of /이/

표 1. 형성음 주파수 검출의 오차수

Table 1. Number of formant detection errors

/아/

|    | $\mu=0.94$ | $\mu=R(1)/R(0)$ | EDP |
|----|------------|-----------------|-----|
| F1 | 0          | 0               | 0   |
| F2 | 19         | 4               | 8   |
| F3 | 21         | 4               | 1   |

/오/

|    | $\mu=0.94$ | $\mu=R(1)/R(0)$ | EDP |
|----|------------|-----------------|-----|
| F1 | 4          | 0               | 2   |
| F2 | 4          | 0               | 1   |
| F3 | 5          | 2               | 1   |

/우/

|    | $\mu=0.94$ | $\mu=R(1)/R(0)$ | EDP |
|----|------------|-----------------|-----|
| F1 | 39         | 23              | 12  |
| F2 | 85         | 84              | 49  |
| F3 | 85         | 82              | 45  |

/에/

|    | $\mu=0.94$ | $\mu=R(1)/R(0)$ | EDP |
|----|------------|-----------------|-----|
| F1 | 2          | 2               | 1   |
| F2 | 3          | 1               | 1   |
| F3 | 4          | 0               | 3   |

| /오/ |            |                 |     |
|-----|------------|-----------------|-----|
|     | $\mu=0.94$ | $\mu=R(1)/R(0)$ | EDP |
| F1  | 6          | 7               | 2   |
| F2  | 62         | 66              | 51  |
| F3  | 61         | 66              | 54  |

F1 : 1st formant

F2 : 2nd formant

F3 : 3rd formant

표 1에서 볼 때, 모음 /아/, /에/, /이/의 경우  $\mu=0.94$  일 경우에 비해서는 EDP방법이 형성음 검출 성능이 훨씬 좋으나,  $\mu=R(1)/R(0)$ 에 비해서는 크게 개선되지 않고 오히려 약간 성능이 안 좋은 듯 하나, /오/, /우/에서는 EDP 방법이 월등히 우수하여, 전체적으로 볼 때 형성음 검출 성능이 개선되었다. 이 결과를 가지고 5개 모음에 대한 형성음 주파수의 검출률을 계산한 결과를 표 2에 실었다.

표 2. 한국어 5개 단모음의 형성음 주파수 검출률  
Table 2. Formant detection rates for 5 Korean vowels

| (단위:%) |            |                 |      |
|--------|------------|-----------------|------|
|        | $\mu=0.94$ | $\mu=R(1)/R(0)$ | EDP  |
| F1     | 95.3       | 97.0            | 98.4 |
| F2     | 84.0       | 85.6            | 89.8 |
| F3     | 83.7       | 85.7            | 90.4 |
| 평균     | 87.7       | 89.4            | 92.9 |

## V. 결 론

본 연구에서는 발음의 시작 부분과 천이 구간에 대하여 성문과에 의한 음원 특성과 입술에서의 방사 특성을 효과적으로 제거함으로써 보다 정확한 분석을 가능하게 하기 위하여 음원 특성과 방사 특성은 근사적으로 음파의 에너지와 비례한다고 가정하고, 정규화된 단구간 에너지를 이용한 에너지 의존 프리엠퍼시스 기법을 제안하였으며, 이를 이용하여 음성 신호 분석 성능을 개선하고자 하였다. 기존에 널리 이용되던 두 가지 프리엠퍼시스에 의하여 같은 실험을 수행하여 성능 비교를 하였다.

얻어진 결론은 다음과 같다. 첫째, 에너지 의존 프리엠퍼

시스를 발음의 시작 부분에서부터 정상 상태까지 적용한 결과, 기존의 방법에 비하여 스펙트럼 형태의 현저한 개선이 이루어졌다. 특히, F1과 F2가 뚜렷이 나타난다. 둘째, 에너지 의존 프리엠퍼시스를 이용할 때, 기존의 방법에 비하여 인접한 두 형성음이 겹쳐서 하나의 형성음처럼 나타나는 현상이 상당히 개선되었다. 셋째, 형성음 주파수의 검출 실험에서 기존의 방법에서 검출이 안되던 위치에서도 에너지 의존 프리엠퍼시스를 적용한 결과, 정확한 형성음을 검출할 수 있었다. 끝으로, 신뢰할만한 주파수 해상도의 FFT의 세 방법의 스펙트럼을 비교한 결과, 발음의 시작 부분에서 천이 구간까지 정확한 공진 꺾으로부터 형성음 주파수가 바이어스(bias)되는 현상이 현저히 개선되었다.

모음의 분석시 발음의 시작 부분을 따로 처리하지 않음으로 인하여 발생할 수 있는 오차를 에너지 의존 프리엠퍼시스를 이용하여 개선할 수 있을 것이며, 또한 자음 분석에서 신호 자체의 음원 특성과 방사 특성은 고정된 것이 아니므로 기존의 고정된 프리엠퍼시스에 비하여 본 연구에서 제안한 방법이 음원과 방사 특성을 더 적절히 제거할 수 있을 것이라 기대된다.


본 연구에서 제안한 방법의 단점중, 식 (9)에서 보듯이, 단구간 에너지의 최대값을 구해야 하는 것인데, 이는 실시간 처리의 장애가 될 수 있다. 따라서 실시간 처리를 위해서는 입력되는 음성 신호의 에너지에 따라 프리엠퍼시스를 실시간으로 조절할 수 있는 방법이 개발되어야 할 것이다.

## 참 고 문 헌

1. J. D. Markel and A. H. Gray, Jr.: *Linear Prediction of Speech*, Springer-Verlag·Berlin·Heidelberg·New York, 1980.
2. S. Saito and K. Nakata: *Fundamentals of Speech Signal Processing*, Academic Press, 1985.
3. J. L. Flanagan: *Speech Analysis Synthesis and Perception* 2nd ed., Springer-Verlag Berlin Heidelberg New York, 1972.
4. J. Orfanidis: *Optimum Signal Processing*, McGraw Hill, 1988.
5. S. L. Orfanidis, "The Double/Direct RLS Lattice," *Proc. ICASSP*, pp. 1381-1382, 1988.
6. 송영수, 안상필, 이재혁, 윤태성, 박상희, "LPC spectrum의 1, 2차 미분을 이용한 Formant 추출," *신호처리 학동학술대회*, 1990. 9.
7. J. I. Makhoul and L. K. Cosell, "Adaptive Lattice Analysis of Speech," *IEEE Trans. ASSP*, Vol. ASSP-29, No. 3, pp. 654-658, June 1981.
8. S. Haykin: *Adaptive Filter Theory*, 2nd ed., Prentice-Hall, 1991.
9. G. Fant: *Acoustic Theory of Speech Production*, Mouton, 1970.
10. R. L. Miller, "Nature of the Vocal Cord Waves," *J. Acoust. Soc. Am.*, Vol. 31, pp. 667-677, 1959.
11. R. B. Mosen, A. M. Engebretson, "Study of Variations in

the Male and Female Glottal Wave," J. Acoust. Soc. Am.,  
Vol. 62, pp. 228-232, 1977.

▲김 등 준(Dong-Jun Kim) 1963년 4월 14일생  
1988년 2월:연세대학교 전기공학과 졸업(공학사)  
1990년 2월:연세대학교 대학원 전기공학과 공학석사)  
1994년 2월:연세대학교 대학원 전기공학과(공학박사)  
1995년 3월~현재:청주대학교 정보통신공학과 전임강사

▲박 상 회(Sang-Hui Park) 1939년 8월 25일생  
 1962년:연세대학교 전기공학과 졸업(공학사)  
1964년:연세대학교 대학원 전기공학과(공학석사)  
1971년:연세대학교 대학원 전기공학과(공학박사)  
1982년:미국 워싱턴대학교 교환교수  
1971년~현재:연세대학교 전기공학과 교수