

포만트 밴드폭 정규화를 이용한 음성인식

正會員 洪宗鎭** 正會員 姜錫建* 正會員 朴健作* 正會員 朴圭泰*

Speech Recognition Using Formant Bandwidth Normalization

Jong Jin HONG,** Suk Kun KANG,* Kun Jahnk PARK,* Kyu Tae PARK* *Regular Members*

要 約 본 논문에서는 기존의 선형예측기법의 문제점을 선형예측계수, AR필터(autoregressive filter)의 pole 위치, 포만트-밴드폭의 관점에서 분석하고, 성문반사계수(glottis reflection coefficient)의 영향을 성도추정이본에 따라 분석했으며, 이러한 분석을 근거로 하여 포만트 밴드폭 정규화방법을 보완하였다. 성문반사계수를 1로 변경하여 성문의 영향을 제거하여 새로운 선형예측계수를 얻을 수 있으며, 이 선형예측계수에 의한 주파수 스펙트럼은 포만트 밴드폭이 0으로 정규화되어 포만트가 최적으로 강조된 스펙트럼이 된다. 이 선형예측계수는 앞뒤로 대칭되면서, 표준편차가 성문반사계수를 변경시키지 않은 선형예측계수보다 증가하므로써, 음성부호화시에 bit rate을 50%로 줄일 수 있으면서 정보의 양을 그대로 보존하고 있음을 알 수 있었다. 이러한 포만트 밴드폭을 0으로 정규화하는 방법을 이용하여 한국어 5개 모음을 포만트에 의해서 소음환경에서 인식하기 위한 실험을 실시하여 96.7%의 인식율을 얻을 수 있었다.

ABSTRACT In this paper, the cause of linear prediction error is analysed and the theoretical basis for normalizing th formant bandwidth to 0 is given and its validity is verified. The formant and bandwidth in relation to the position of the poles of AR filter are measured for an analysis of the relation between the pole position and the formant bandwidth. By changing the glottis reflection coefficient to 1, the effect of the glottis is eliminated and as the result a new linear prediction coefficients are obtained by normalizing the formant bandwidth of the signal to 0. Since these coefficients are symmetrical, the standard deviation is larger than the coefficients with fixed glottis reflection coefficient. The bit rate for speech coding can be reduced by a factor of 2 without any loss of information. Through computer simulation, recognition rate of 96.7% is obtained by using the proposed algorithm in recognizing 5 Korean vowels in noisy environment.

I. 서 론

음성신호를 인식, 저장 및 전송하기 위해서는 음성 신호를 부호화 해야 하는데 이러한 부호화에서의 주요 문제는 음성신호의 중복성을 제거하면서 음성의 특징을 잘 나타내는 파라메타를 찾아내어 부호화하는 것이다 이러한 파라메타에는 시간영역 파라메타와 주파수영역 파라메타가

있다.

주파수영역 파라메타에는 주로 선형예측계수(LPC, Linear Prediction Coefficients), 반사계수(Reflection Coefficients), 포만트 주파수(Formant Frequency), 성도면적함수(vocal tract area function)가 있다.

주파수영역 파라메타의 관계는 그림 1과 같이 나타낼 수 있는데 그림에서 선형예측계수에 의한 다항식을 인수분해하여 z-평면에서 pole의 위치를 구하여 포만트와 밴드폭을 구한 것과, 선형예측계수를 바로 푸리에 변환하여 포만트와 밴드폭을 구한 것과는 차이가 있다. 따라서 선형예측계

*延世大學校 電子工學科
Dept. of Electronics, Yonsei University
**陸軍3775부대 1포대장.
論文番號 : 91-44 (接受1991. 2. 5)

수와 포만트, 밴드폭의 관계 규명이 필요하다. 그리고 z -평면에서의 pole의 위치와 포만트 및 밴드폭의 관계는 그림 1에서 나타낸 식의 관계가 성립하는 것으로 지금까지 사용하여 왔으나 실제로는 이러한 관계가 항상 성립하는 것이 아니고 pole의 위치가 단위원에 가까이 갈 때에만 이런 관계가 성립하는 것을 실제 계산을 통하여 알 수 있다.

여기서 F_m 은 포만트 주파수, B_w 는 포만트 밴드폭, R 은 2차 공진기의 근(pole)의 z domain에서 원점으로부터의 거리, θ 는 근의 양의 실축으로부터의 거리, h_k 는 inverse filter의 계수 즉 선형예측계수이고, T 는 sampling 주기, f_s 는 sampling 주파수이다.

$$H_{AR}(z) = \frac{1}{1+h_1z^{-1}+h_2z^{-2}} \quad (식1)$$

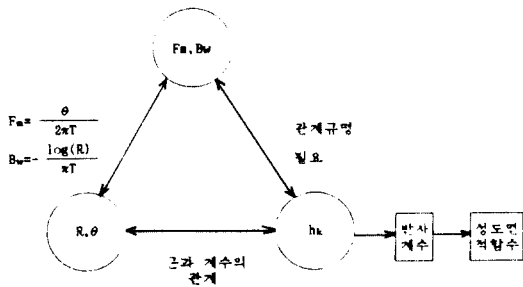


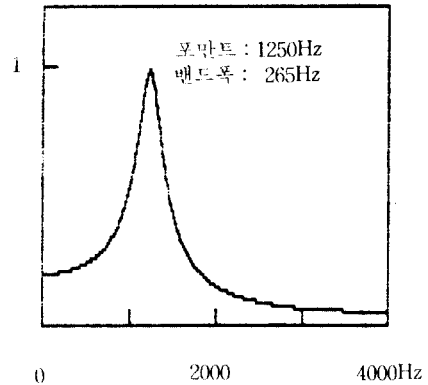
그림 1. $F_m, B_w-R<\theta-h_k$ 간의 관계
Fig. 1. Relations between F_m, B_w and $R<\theta$ and h_k

또 Honda와 Itakura는 1984년 선형예측계수에 동일한 가중치를 부여하여 음성을 부호화 했으나 [1], 실제로 (식1)과 같은 선형예측계수에 의한 2차필터의 계수값을 변화시켜가면서 스펙트럼을 구해본 결과 그림 2와 같이 얻을 수 있었다. 그림 2에서 (a)는 h_1 이 -1.0 , h_2 가 0.8 , (b)는 h_1 이 -1.5 , h_2 가 0.8 , (c)는 h_1 이 -1.0 , h_2 가 0.3 일때의 스펙트럼이다. 이 그림에서 알 수 있는 바와 같이 (b)와 같이 선형예측계수 h_1 이 변화할 때는 밴드폭의 변화없이 포만트만 변화했으나,

(c)와 같이 h_2 가 변화할 때는 포만트와 밴드폭이 동시에 변화하는 것을 알 수 있다. 따라서 각각의 선형예측계수에 동일한 가중치를 부여하는 것은 잘못된 것임을 알 수 있다.

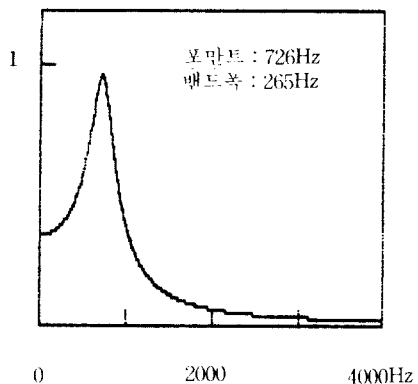
본 논문에서는 포만트-밴드폭과 선형예측계수와의 관계를 정립하고 선형예측계수에 가중치를 주는 문제의 기초 연구를 하며, 성도추정이론에 근거하여 선형예측계수의 오차를 분석하고, 그에 대한 대안으로써 포만트밴드폭 정규화 방법 [2]의 이론적 타당성을 분석하고 이 방법을 실제 음성신호의 인식실험에 적용하여 그 타당성을 입증하였다.

크기

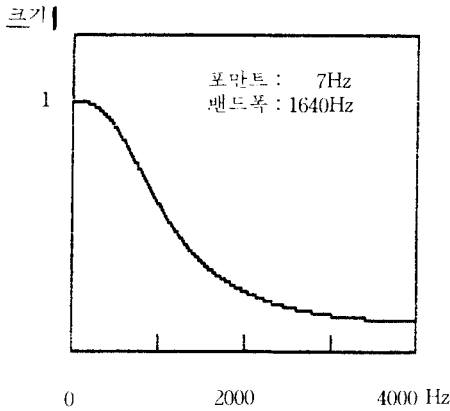


(a) $h_1=-1.0, h_2=0.8$

크기



(b) $h_1=-1.5, h_2=0.8$



(c) $h_1=-1.0, h_2=0.3$

그림 2. 선형예측계수의 변화에 따른 스펙트럼의 변화
Fig. 2. change in spectrum according to the variation of LPC

II. 파라메타간의 관계

1. F_m 과 L_θ 의 관계

그림 3과 같이 Z-domain 상에 pole이 존재하는 2차 AR 필터의 전달함수는 다음과 같이 구할 수 있다. 즉

$$H_{AR}(Z) = \frac{k}{1 - 2R\cos\theta z^{-1} + R^2 z^{-2}} \quad (식2)$$

이 된다. 여기에서 F_m 과 θ 의 관계는

$$F_m = \frac{\theta}{2\pi} f_s \quad (식3)$$

의 관계가 성립하는 것으로 알고 있다[3]. 여기서 sampling frequency는 8kHz이다. 그러나 실제로 (식2)에서 θ 을 0에서 10도씩 변화시켜가면서 전달함수의 스펙트럼을 구하여 포맷트를 구해보면 θ 에 대한 F_m 의 관계는 그림 5와 같음을 알 수 있다. 즉 R의 값에 따라서 포맷트가 달라진다. 여기서 R값이 1에 가까워질 때는 (식3)의 관계가 성립한다는 것을 알 수 있다.

2. B_w 와 R의 관계

R과 B_w 의 관계는

$$B_w = \frac{\log(R)}{\pi T} \quad (식4)$$

가 성립하는 것으로 알고 있으나[3] 위의 방법과 같이 (식2)에서 각각 θ 에 대하여 R을 0에서 0.1씩 증가시켜가면서 밴드폭(B_w)을 측정해 보면 그림 6의 결과가 나온다. 즉 B_w 는 θ 별로 각각의 R값에 따라서 다르다. R값이 1이면 밴드폭은 0이다.

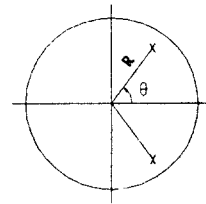


그림 3. Z-domain에서 pole의 위치
Fig. 3. pole location on z-plane

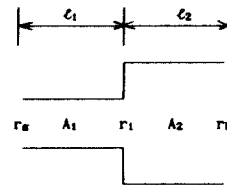


그림 4. 2차 공진기
Fig. 4. 2th order lossless resonator

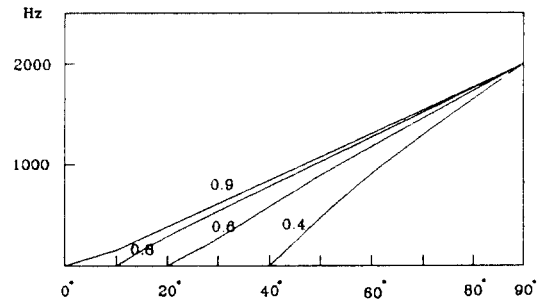


그림 5. θ 의 변화에 따른 포맷트의 변화
Fig. 5. Change in formant according to the variation of θ

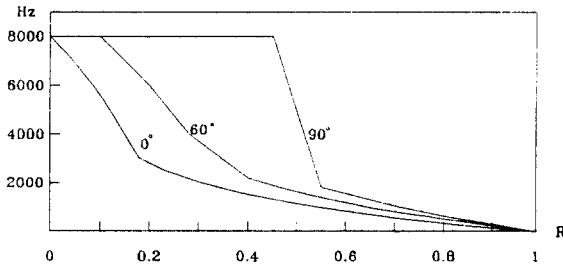


그림 6. R의 변화에 따른 밴드폭의 변화
 Fig. 6. Change in bandwidth according to the variation of R

Ⅲ. 선형예측기법에 의한 포만트 분석시 문제점

1. 2차 튜브 분석

그림 4의 2차 공진기는 사람의 성도를 무손실 튜브로 모델링한 것이다. 그림에서 단면적 A_1 과 A_2 는 해당 section에서 일정하며, 길이 l_1 과 l_2 는 같다. 여기서 r_g 는 성문반사계수, r_L 은 입술에서의 반사계수, r_1 은 section의 접합지점에서의 반사계수이다. 이 2차 유향튜브의 volume velocity 에 의한 전달함수는 다음과 같다[3].

$$H_{vt}(z) = \frac{k}{1 + (r_g + r_L)r_1 z^{-1} + r_g r_L z^{-2}} \quad (식5)$$

이 때 z 는 음파가 각 section을 왕복하는데 걸리는 시간이다. 이 왕복시간은 유향신호를 획득할 때 사용되는 sampling 주기이다. 이 전달함수는 2차 AR(auto-regressive) 시스템의 전달함수와 형태가 동일하다. 따라서 이 시스템의 정확한 인버스 필터를 선형예측기법으로 부터 얻어낼 수가 있다. 또 인버스 필터의 역으로부터 시스템의 AR 모델을 다음과 같이 얻을 수 있다.

$$H_{AR}(z) = \frac{k}{1 + h_1 z^{-1} + h_2 z^{-2}} \quad (식6)$$

여기서 h_1 과 h_2 는 선형예측기의 계수이다. 상기

(식5)와 (식6)에서 선형예측계수와 반수계수의 관계를 발견할 수 있다. 즉 예측계수 h_2 입술과 성문에서의 손실에 관계가 있는데 반하여 h_1 는 그에 추가하여 성도의 모양과 관계가 있는 반사계수 r_1 과 관계가 있다. 따라서 그림 2에서 나타난 바와 같이 h_1 이 변화할 때는 스펙트럼의 포만트가 변화하고 h_2 가 변화하면 포만트와 밴드폭이 같이 변화하는 이유를 이해할 수 있다.

또 선형예측기법으로 부터는 h_1, h_2 2개의 계수 값을 얻을 수 있는데 반하여 반사계수는 r_g, r_L, r_1 3개이다. 따라서 선형예측기법으로 부터 성도추정을 위하여 반사계수를 구하고자 할 때는 일반적으로 r_g 또는 r_L 을 1이라고 가정을 하고 계산을 해왔다[5]. 이 가정은 자유공간의 크기가 입술의 크기보다 대단히 크다고 보았을 때, 입술에서의 반사계수 r_L 이 거의 1에 가까운 값을 갖는다고 보거나 음성생성시 성문이 폐쇄되어 있어서 완벽한 반사가 성문에서 일어나기 때문에 성문반사계수를 1로 가정한 것이다. Wakita의 방식대로 r_L 을 1이라고 가정하면 상기 5, 6식으로 부터 $r_g = h_2$ 가 되며, 이는 r_g 가 2차 공진기의 공진 대역폭을 나타낸다는 것을 보여준다. 또 r_g 값의 변화는 필터계수 h_1 에 변화를 일으켜서 공진 주파수를 변화시킨다. 이것은 성문의 반사계수에 따라서 비록 성도의 모양이 변하지 않더라도 포만트주파수와 대역폭이 변화할 수 있다는 것을 보여주는 것으로써, 기존에 사용된 포만트로는 즉, 음소는 포만트에 의해서 특징 지을 수 있다는 이론이 잘못되었음을 나타낸다. 즉 기존의 선형예측기법에 의한 선형예측계수는 성문반사계수에 의한 영향을 포함하고 있기 때문에, 이 예측계수에 의한 스펙트럼은 그림7과 같이 실제 성도의 모양을 바르게 나타내지 못하고 있다.

따라서 동일한 성도의 모양으로부터 r_g 의 값에 따라서 거의 무한대의 각기 다른 음파가 출력될 수 있으며, 이것이 동일한 화자가 성도를 변화시키지 않아도 포만트의 변이가 생긴 이유으로써 기존의 음성인식에 오차를 유발하는 원인중의 하나였다. 유향 튜브의 모양은 r_1 만이 결정될

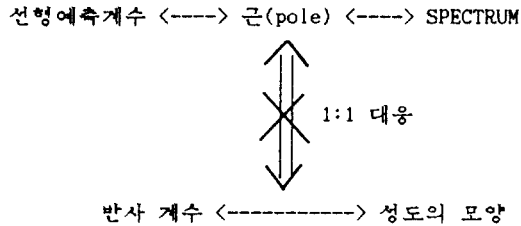


그림 7. 선형예측계수와 성도의 모양과의 관계
 Fig. 7. Relation between LPC and vocal tract shape

하고 r_g 의 값은 실제 튜브의 모양에 영향을 미치지 않는다. 따라서 선형예측계수는 그 근이 포매파트와 대역폭을 나타내고 반사계수는 성도의 모양을 보다 직접적으로 나타낸다. 그러나 포매파트와 대역폭은 성문반사계수에 의해서 그 값이 변하지만 성도의 모양은 성문반사계수와 무관하다. 따라서 포매파트와 대역폭은 어떤 값으로 정규화시켜야 시스템과 일대일로 대응시킬 수 있다.

2. 성문반사계수의 변화에 따른 전달함수의 변화

성문반사계수 r_g 의 변화에 따라서 시스템의 전달함수가 변화하고 이에 따라서 pole의 위치가 변하며, 포매파트와 대역폭이 변한다는 사실을 알아보기 위해서 (식5)에서 r_L 을 1로 놓고, r_g 를 0에서 1까지 0.1씩 증가시키면서 z plane 위에 근의 위치를 표시하면 그림 8과 같이 근의 위치가 변한다. 그림 8은 반사계수 r_L 이 -0.6, 0.6일 때의 경우이다.

그림 8에서 볼 수 있는 바와 같이 r_g 의 변화에 대하여 2차 유향 튜브의 근의 궤적은 비선형이 되어, 그림 5, 6에서 나타난 바와 같이 동일한 반사계수 r_L 에 대하여 해당 유향 튜브의 포매파트와 밴드폭은 동시에 변화한다. 이렇게 근의 궤적이 비선형이 되는 이유를 분석해 보면 (식5)에서 이 전달함수의 근의 위치를 구하기 위하여 분자 분모항에 z^2 을 곱하고 $r_L=1$ 이면

$$H_{vt}(z) = \frac{kz^2}{z^2 + (r_g + 1)r_L z + r_g} \quad (식7)$$

가 된다. (식7)의 분모의 근은 a, b 라 하고, 근의 실수분은 x , 허수분은 y 라 하면 $a, b = x \pm jy$ 라 할 수 있다. 2차 방정식의 근의 공식에 따라

$$x = \frac{-(r_g + 1)r_L}{2} \quad (식8)$$

$$y = \frac{\sqrt{(r_g + 1)^2 r_L^2 - 4r_g}}{2} = 0.5(r_g r_L - 2\sqrt{r_g} + r_L) \quad (식9)$$

(식8)을 r_g 에 대하여 풀고 x 에 대하여 미분하고, (식9)을 r_g 에 관해 미분하면

$$\frac{dr_g}{dx} = -\frac{2}{r_L} \quad \frac{dy}{dr_g} = 0.5\left(r_L - \frac{1}{\sqrt{r_g}}\right) \quad (식10)$$

따라서 x 의 변화에 대한 y 의 변화를 구해보면

$$\frac{dy}{dx} = \frac{dy}{dr_g} \times \frac{dr_g}{dx} = \left(\frac{1}{r_L \sqrt{r_g}} - 1\right) \quad (식11)$$

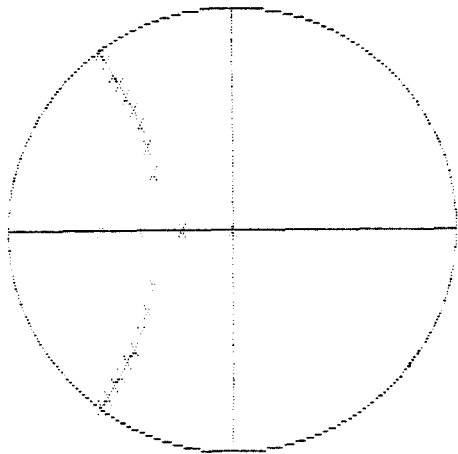
(식11)에서 실수분의 변화에 대한 허수분의 변화는 정수가 아니고 $\sqrt{r_g}$ 에 반비례하기 때문에 z plane에서 r_g 의 변화에 대한 근의 궤적은 곡선을 그린다. 또 r_g 가 1이 되면 식(7)은

$$H_{vt}(z) = \frac{kz^2}{z^2 + 2r_L z + 1}$$

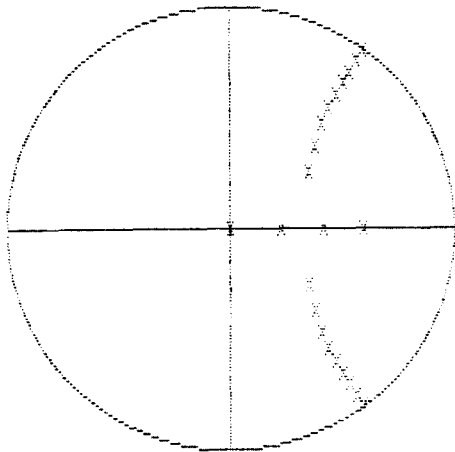
이 되고 근 $a, b = -r_L \pm j\sqrt{1 - r_L^2}$ 이 된다. 또 원점에서의 거리 R 은

$$R = \sqrt{(-r_L)^2 + (\sqrt{1 - r_L^2})^2} = 1$$

따라서 근은 unit circle 위로 이동하여 그림 6의 결과와 같이 밴드폭은 0이 된다.



(a) $r_1 = 0.6$ 일 때



(b) $r_1 = -0.6$ 일 때

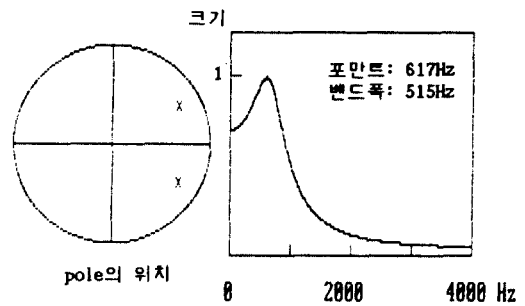
그림 8. 상분반사계수의 변화에 따른 pole 위치의 변화
Fig. 8. Change in pole location according to the variation of LPC

3. 인접 pole에 따른 포먼트와 밴드폭의 변화

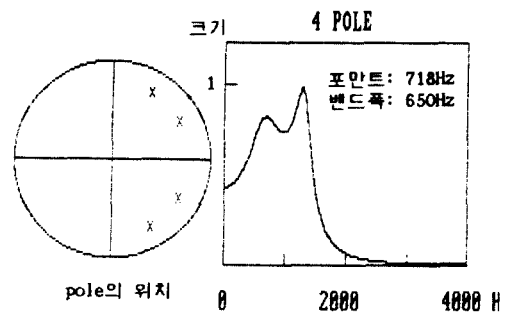
그림 9에서 알 수 있는 바와 같이 (a)의 2pole 필터의 경우 스펙트럼의 포먼트가 617 Hz, 밴드폭

이 515Hz이었으나, (b)와 같이 인접해서 다른 pole이 존재할 때는 2pole일 때의 포먼트와 밴드폭이 포먼트는 718Hz, 밴드폭은 650Hz로 변화하는 것을 알 수 있다.

따라서 선형예측계수에는 인접한 pole의 영향으로 인한 오차를 함께 포함하고 있음을 알 수 있으며, 이것은 그림 1을 이용하여 설명한 바와 같이 선형예측계수에 의한 다항식을 인수분해하여 pole의 위치를 구하여 포먼트와 밴드폭을 구한 것과 선형예측계수를 직접 푸리에 변환하여 스펙트럼을 구하여 포먼트와 밴드폭을 구한 것이 차이가 나는 이유의 하나가 된다.



(a) 2 pole일 때



(b) 4 pole 일 때

그림 9. pole의 위치에 따른 포먼트와 밴드폭 변화
Fig. 9. Change in formant and bandwidth according to the pole location.

4. 선형예측계수 오차의 의미

이 상에서 알아본 바와 같이 선형예측계수에 의한 오차는 다음과 같은 오차를 포함하고 있음

을 알 수 있다.

- 성문반사계수에 의한 오차
- 인접한 pole의 영향에 의한 오차
- 입술에서의 반사계수의 가절치에 의한 오차

위와 같은 오차에서 성문반사계수에 의한 오차와 인접한 pole의 영향에 의한 오차는, 성문반사계수를 1로 변환하여 성도 전달함수의 pole의 위치를 단위원 위로 이동시켜서 포만트 밴드폭을 0으로 정규화함으로써 제거할 수 있으며, 선형예측계수와 성도의 모양을 1:1로 대응시킬 수 있다.

IV. 포만트 밴드폭 정규화

1. step-up 과정

선형예측계수에서 성문반사계수에 의한 영향을 없애고 오차를 최소화 하며, 예측계수와 시스템을 일대일로 대응시키기 위해서는 선형예측계수를 구한 후에 반사계수를 구한 다음, 반사계수 중에서 성문반사계수 r_g 를 1로 바꾼 후에, 이 반사계수에 의한 새로운 선형예측계수를 구해야 한다. 반사계수에서 선형예측계수를 구하는 과정을 step-up[4]이라 한다. 아래 식에서 보는 바와 같이 m 차의 선형예측계수를 구할 때 다음의 식에 의하여 예측계수를 구할 수 있다.

$$\begin{aligned}
 a_{m1} &= a_{m-1, 1} \quad (i=0 \text{ 일 때}) \\
 a_{m-1, 1} + k_m a_{m-1, m-1} & \quad (i=1, 2, \dots, m-1 \text{ 일 때}) \\
 k_m & \quad (i=m \text{ 일 때})
 \end{aligned} \tag{식12}$$

2. 포만트 밴드폭 정규화의 장점

위의 step-up 절차에 의하여 성문반사계수를 1로 바꿈으로써, 성문반사계수의 영향을 제거하여 얻은 새로운 선형예측계수에 의해서 스펙트럼을 얻으면, 성문반사계수의 영향을 제거한 스펙트럼을 얻을 수 있다. 그림 10에는 '아'음에 대하여, 표 1의 선형예측계수 즉, 본래의 선형예측계

수와 step-up과정을 통하여 포만트밴드폭 정규화에 의해서 얻은 선형예측계수에 의한 스펙트럼을 비교해 놓았다.

그림 10에서 알 수 있는 바와 같이 성문반사계수를 1로 바꾸어 얻은 선형예측계수의 스펙트럼은 포만트 밴드폭을 0으로 바꾸어 줌으로써, 본래의 성문반사계수에 의한 영향을 포함하고 있는 선형예측계수에 의한 스펙트럼에서는 나타나지 않는 peak를 잘 나타내는 것을 알 수 있으며, 인접한 pole에 의한 영향이 제거된다는 것을 알 수 있다. 그러므로 포만트 밴드폭을 0으로 정규화한 스펙트럼에서 peak 즉 포만트를 찾을 때는 간단한 peak-picking 방법에 의해서 보다 신속하게 포만트를 찾을 수 있으며, 그림 9에서 나타나는 인접 pole에 의한 영향을 제거한 포만트를 얻을 수 있다.

또 선형예측계수의 오차 발생시에는 포만트와 밴드폭의 오차가 동시에 발생하지만, 포만트 밴드폭을 0으로 정규화함으로써 포만트와 밴드폭의 2차원 오차를 포만트만의 1차원 오차로 바꾸어 포만트에 의한 음성인식에 인식시간을 보다 단축할 수 있다.

반사계수에 의한 음성인식시에는 성문반사계수를 비교 대상으로 제외함으로써 인식시간을 단축할 수 있다.

성문반사계수를 1로 바꾸어 얻은 선형예측계수는 처음과 끝의 계수가 서로 대칭된 계수를 얻을 수 있어서 음성 전송을 위해 코딩(coding)시에 bit rate를 50%로 감소시킬 수가 있으며, 선형예측계수에 의한 음성인식시에도 1/2의 계수만 비교하면 되므로 인식시간을 단축할 수 있다.

표 1에는 '아'음에 대하여 autocorrelation 방법으로 얻은 8차의 선형예측계수를 비교해 놓았다. 선형예측계수가 대칭이 되어 코딩의 bit rate를 줄일 수 있으며, 그림 11에서 볼 수 있는 바와 같이 포만트 밴드폭을 정규화한 경우의 선형예측계수의 표준편차가 기존의 선형예측계수의 표준편차보다 증가됨을 알 수 있다. 그림 11은 40명의 '아'음의 선형예측계수의 크기와 표준편차를 나타낸 것이다. 이 그림에서 알 수

있는 것은, 포맷트 밴드폭 정규화 방법에 의하여 선형예측계수가 대칭이 되므로써 코딩비율을 50%로 줄일 수 있으며, 대칭된 선형예측계수의 표준편차가 증가함으로써 실제 정보량은 상실된 것이 아니라 그대로 보존되고 있다는 것이다.

선형예측계수	Original $r_g=0.2$	Normalized $r_g=0.9999$
$h(0)$	1.0	1.0
$h(1)$	-0.15798	-0.42414
$h(2)$	-0.26918	-0.48131
$h(3)$	0.195715	0.680618
$h(4)$	0.176686	0.294678
$h(5)$	0.620058	0.680818
$h(6)$	-0.30905	-0.48131
$h(7)$	-0.30159	-0.42414
$h(8)$	0.2	0.99999

표 1. 선형예측계수의 비교
Table. 1. Comparison between LPCs

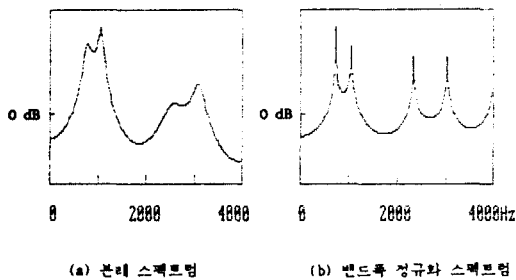
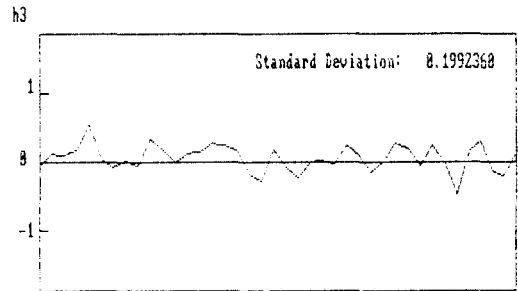


그림 10. 표 1의 선형예측계수에 의한 스펙트럼의 비교
Fig. 1. Comparison between spectrum by LPCs in table 1.

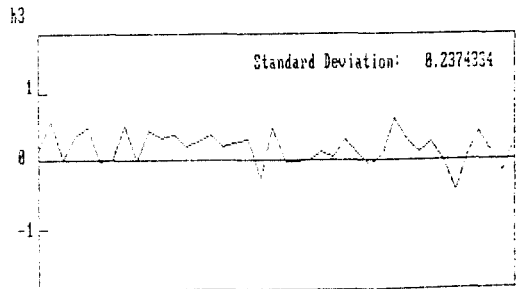
V. 실험 및 결과 고찰

1. 실험 방법

본 논문에서는 포맷트 밴드폭 정규화방법이 음성의 인식에 어떠한 결과를 나타내는가를 알기 위하여 한국어 ‘아’, ‘에’, ‘이’, ‘오’, ‘우’ 5개음에 대해서만 포맷트에 의한 인식실험을 했다.



(a) 본래의 예측계수



(b) 밴드폭 정규화에 의한 예측계수

그림 11. 선형예측계수의 표준 편차
Fig. 11. Standard deviation of LPCs

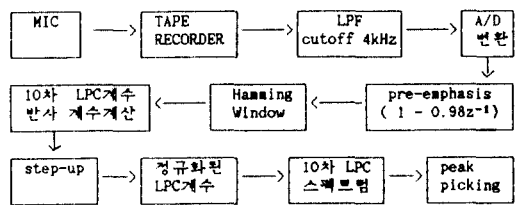


그림 12. 실험 블록 다이어그램
Fig. 12. block diagram of experiments

그림 12에는 실험의 블록 다이어그램이 나와있다. 음성데이터는 먼저 20대의 남성 10명의 음성을 cassette tape recorder에 녹음하여 차단주파수가 4kHz인 저역통과 필터를 통과시킨 후에 DT 2821 16SE A/D, D/A 보드와 ILS(Interactive Laboratory System) 신호처리 소프트웨어를 이용해서, 8kHz 샘플링과 12bit의 resolution으로 데이터를 만들었다. 본 실험에서는 소음환경에서

의 인식율을 구하기 위하여 음성 녹음시에 일반 가정용 카세트 테이프레코더를 사용하므로써 녹음기 자체의 기계적 잡음을 입력했으며, 실험실에서 사람들이 자연스럽게 대화를 하는 환경에서 녹음을 했다. 그 결과 신호대잡음비(SNR)가 14~24dB가 되며 평균 19dB인 상황에서, 음성 인식을 위한 템플레이트(template)를 만들기 위하여 먼저 10명의 '아', '에', '이', '오', '우' 음성에 대하여 각각 1024 포인트를 잘라내어 1 frame 데이터를 만들었다. 이렇게 1 frame의 길이를 1024 point로 길게 설정한 이유는 음성신호에 잡음이 많이 섞여 있으므로 1개의 frame에 여러개의 음성 피치주기가 포함되게 함으로써 잡음을 평균하여 잡음의 효과를 줄이기 위한 것이다.

이 데이터의 주파수 스펙트럼상의 dynamic range를 줄이기 위하여 $1-0.98z^{-1}$ 의 pre-emphasis 필터를 통과시켰다. 또 이 데이터를 hamming window에 통과시켜서, 1024 points의 frame을 만들면서 발생하는 window의 edge 효과를 줄였다. 이 데이터를 이용하여 autocorrelation 방법으로 10차 선형예측계수를 구한 다음 반사계수를 구하여, 이 반사계수중에서 성분반사계수 r_g 를 0.9999로 바꾼 후에 step up [4] 방법으로 새로운 선형예측계수를 구한 후에, 이 선형예측계수를 이용하여 1024 포인트 FFT를 취하여 스펙트럼을 구한 다음 F1, F2, F3 포먼트를 구했다. 이 포먼트를 10명에 대하여 평균하여 각 음성별 template로 사용했다. 다음 같은 방법으로 새로운 30명의 음성에 대하여 데이터를 만들어서 각각 3개의 포먼트를 구한 다음, 각 음성의 3개 포먼트와 template와의 distance를 계산하여 이 distance가 가장 작은 음을 인식한 것으로 했다.

2. 5개 모음 인식 결과

20대의 성인 남자 30명이 5개음을 1회씩 발음한 150개음에 대하여 앞에서 설명한 방법에 의해서 실험한 결과는 다음 표와 같다.

음성 방법	음성					평균
	아	에	이	오	우	
기존방법	86.7	86.7	86.7	33.3	3	59.2
포먼트 정규화	100	100	100	86.7	96.6	96.7

표 2. 한국어 5개 모음에 대한 인식율
Table 2. Recognition rate of 5 Korean vowels

이 실험의 결과에서 알 수 있는 바와 같이 포먼트밴드폭 정규화 방법은 소음환경에서 양호하게 음성인식을 하는 것을 알 수 있다.

VI. 결 론

본 논문에서는 기존 선형예측기법의 문제점을 상도추정이론 및 선형예측계수, pole의 위치, 포먼트-밴드폭의 관점에서 분석하고, 이에 근거하여 포먼트 밴드폭 정규화 방법[2]을 보완하여 제시했다. 포먼트 밴드폭을 0으로 정규화하기 위하여 성분반사계수를 1로 바꾸어 줌으로써 포먼트-밴드폭의 2차원 오차가 포먼트 반의 1차원 오차로 변경되고, 선형예측계수와 상도의 모양간에 1:1 대응관계가 성립되며, 성분반사계수의 영향을 제거한 최적 스펙트럼을 얻을 수 있다. 포먼트 밴드폭 정규화가 실제음성의 인식에 미치는 영향을 알기 위하여 포먼트 밴드폭 정규화를 이용하여 5개의 한국어 단모음 인식 실험을 했다. 소음환경에서 10명의 20대 성인 남성의 음성데이터로 보통의 선형예측기법 및 포먼트밴드폭 정규화에 의한 포먼트 template를 만들고, 이를 이용하여 30명의 음성에 대해 인식을 시도하여 96.7%의 인식율을 얻었다. 또 선형예측계수에 동일한 가중치를 주어 코딩을 하는 것이 잘못임을 스펙트럼의 분석을 통하여 나타냈다.

앞으로의 과제는 2차 이상의 선형예측계수의 계수별 가중치 부여 문제와 선형예측계수의 오차와 반사계수 및 상도면적함수의 오차관계를 분석

하는 문제와 전체 파라메타간의 과제를 고려한 최적 파라메타 분석등의 과제가 있다. 본 논문에서는 포먼트 밴드폭 정규화 방법을 한국어 5개 모음에 대해서 인식 실험을 했으나, 차후로는 소음이 없는 환경을 비롯하여 여러가지 환경속에서 다른 모음및자음, 그리고 연결음에 대해서도 이 방법을 사용하여 그 결과를 관찰하는 것이 필요하다고 생각된다.

參 考 文 獻

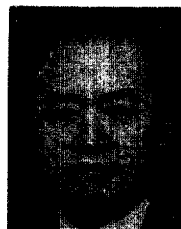
1. M. Honda, F. Itakura, Bit Allocation in Time and Frequency Domains for Predictive Coding of Speech,

IEEE Trans, ASSP Vol. 32, No. 3, pp. 465~473, June, 1984.
2. 한희, 성도의 변화 추정을 위한 시변선형예측기법, 박사 학위 논문, 1989.
3. L.R. Rabiner and R.W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1978.
4. J.D. Markel, Linear Prediction of Speech, Springer-Verlag, Berlin, Heidelberg, New York, 1976.
5. H. Wakita, Direct Estimation of the Vocal Tract Shape by Inverse Filtering of Acoustic Speech Waveform, IEEE Trans, ASSP, Vol. AU 21, pp. 617~627, 1973.

洪宗鎭(Jong Jin HONG) 正會員
1982年 3月 : 陸士 卒業
1991年 2月 : 延世大 電子工學科 本大學院 卒業(碩士)
1991年 6月 現在 : 陸軍3775部隊 1 副隊長

姜錫建(Suk Kun KANG) 正會員
1981年 2月 : 延世大 電子科 卒業
1986年 2月 : 延世大 電子科 本大學院 卒業(碩士)
1987年 6月 ~ 現在 : 延世大 電子科 博士課程 在學

朴健作(Kun Jahk PARK) 正會員
1968年 3月 : 仁荷大 電氣工學科 卒業
1970年 8月 : 서울大 電子工學科 本大學院 卒業(碩士)
1982年 6月 ~ 現在 : 延世大 電子工學科 博士課程
現在 : 昌原技能大學 副教授



朴圭泰(Kyu Tae PARK) 正會員
1933年 6月 11日生
1953年 ~ 1957年 : 延世大學校電氣工學科 卒業
1957年 ~ 1964年 : 延世大學校 本大學院 電子工學科 卒業(工學碩士)
1962年 ~ 1964年 : 英國 LONDON UNIVERSITY, MSc(工學碩士)
1967年 ~ 1969年 : 英國 SOUTHAMPTON UNIVERSITY, Ph. D. (工學博士)
1970年 ~ 現在 : 延世大學校 教授