

## 유색 잡음 환경하에서 Cumulant를 이용한 한국어 단모음 인식

### Korean Single-Vowel Recognition Using Cumulants in Color Noisy Environment

이형근\*, 양원영\*, 조용수\*\*

(Hyung-Gun Lee\*, Won-Young Yang\* and Yong-Soo Cho\*\*)

#### 요약

본 논문에서는 3차 cumulant를 이용하여 음성의 특징벡터를 추출하고, 이것을 신경회로망의 입력으로 사용하는 음성 인식 방법을 제시한다. 3차 이상의 고차 cumulant를 이용하면 Gaussian 잡음과 음성 신호의 분리가 가능하며, 충분히 많은 데이터를 사용할 경우 음성 데이터를 AR 모델링한 계수값을 bias 없이 추출할 수 있다. 또한 기존의 2차 statistics를 이용한 특징 벡터 추출 방법과 비교할 때 잡음이 큰 경우에도 분산은 크지만 bias가 작아 보다 잡음에 강한 특징벡터를 추출할 수 있다. 한국어 단모음에 대한 모의실험을 통하여 유색 잡음 환경에서 SNR이 커질수록 3차 cumulant를 이용한 방법이 기존의 2차 statistics를 이용한 방법보다 높은 인식율을 나타냄을 보인다.

#### ABSTRACT

This paper presents a speech recognition method utilizing third-order cumulants as a feature vector and a neural network for recognition. The use of higher-order cumulants provides desirable uncoupling between the gaussian noise and speech, which enables us to estimate the coefficients of AR model without bias. Unlike the conventional method using second-order statistics, the proposed one exhibits low bias even in SNR as low as 0 dB at the expense of higher variance. It is confirmed through computer simulation that recognition rate of korean single-vowels with the cumulant-based method is much higher than the results with the conventional method even in low SNR.

#### 1. 서론

음성 인식 기술이 최근 10여년간 급속히 발달하여 고립 단어 인식용 chip을 시장에서 손쉽게 구입할 수

있으며 연속 단어 인식 시스템의 개발이 연구실에서 한창 진행 중이다. 이러한 음성 인식 시스템의 성능을 향상시키기 위해서는 DTW, HMM, 신경망 등의 인식 알고리즘과 입력 단에 사용되는 특징 파라미터의 적절한 선택이 필요하다. 특히 이 특징 벡터를 설계할 때, 첨가되는 잡음의 영향을 적게 받아 실제 상황에서 인식률의 저하를 방지하는 방식이 필요하나 현재까지 발표된 대부분의 특징 벡터는 잡음에 대한

\*중앙대학교 공과대학 전기공학과  
Department of Electrical Engineering

\*\*중앙대학교 공과대학 전자공학과  
Department of Electronic Engineering Chung-Ang  
University

접수일자: 1993년 11월 12일

면역성이 크지 않다. 즉, 비교적 잡음이 적거나 조용한 연구실 내에서는 잘 작동하는 시스템이 달리는 자동차 내에서처럼 주변 잡음이 비교적 크고 잡음들 사이에 상관관계가 존재하는 실제 상황에서는 일반적으로 인식률이 급격히 떨어지는 것을 알 수 있다. 예를 들면, autocorrelation 방법이나 covariance 방법을 이용하여 추정되는 AR 계수를 특징 벡터로 사용할 때, 이들은 백색 잡음환경 하에서 동작한다는 가정 하에서 유도되므로 실제 잡음에 가까운 유색 잡음 환경 하에서는 급격한 성능 저하를 볼 수 있다.

최근 고차 스펙트럼에 대한 연구가 활발하게 진행되면서 음성 신호 처리 분야에서도 고차 스펙트럼을 이용하고자 하는 연구가 진행되어 왔다. 또한 컴퓨터와 신호 처리 기술의 발달로 고차 statistics에 대한 연구가 활발히 진행되어 현재 이에 대한 중요성 성질들과 추정 방법들이 최근 많은 문헌에 발표되었다<sup>[1][2][3][4]</sup>. 고차 스펙트럼은 비최소 위상 신호의 재생, Gaussian 잡음의 제거, 그리고 시스템의 비선형성을 추정하기 위하여 주로 사용된다. 음성 인식의 경우 유색 Gaussian 잡음을 제거하기 위하여 3차 이상의 고차 cumulant가 0이되는 성질을 이용한다. 이는 Gaussian 또는 대칭성의 확률 분포 함수를 갖는 잡음과 음성신호의 모음과 같은 non-Gaussian 특성을 갖는 신호의 cumulant가 독립적이고 잡음의 cumulant가 0이 되기 때문이다.

최근까지 고차 statistics에 대한 기본적 성질과 추정방법을 제시하는 많은 논문이 발표되었으며, 2차 statistics를 사용하여 구한 특징 벡터를 입력으로 하는 음성인식 방법들이 제안되었다. 본 논문에서는 유색 잡음에 대한 면역성이 강한 3차 cumulant<sup>[1][3][4][5][6]</sup>를 이용한 한국어 단모음 인식 방법을 제안한다. 인식 방법으로는 3차 cumulant를 이용하여 구한 AR 계수 값을 입력으로 하는 BP(Back Propagation) 신경 회로망을 사용한다. 모의실험 분석을 위하여 각기 SNR이 다른 유색 잡음 환경에서의 7개 한국어 합성 모음(/아/, /에/, /이/, /오/, /우/, /으/, /어/)을 사용하였다. 먼저 신경 회로망의 입력 단에 사용될 특징 벡터를 형성하기 위하여 잡음의 크기를 변화시키면서 기존의 방법과 3차 cumulant를 이용한 방법으로 추정된 AR 계수를 합성시 사용된 계수 값과 비교해 보고 데이터의 갯수에 따른 variance 및 bias를 비교해 보았다. 또한 BP 신경 회로망을 이용한 음성 인식

실험을 통하여 3차 cumulant를 이용한 방법이 기존의 2차 statistics을 이용한 방법보다 우수함을 입증한다.

## II. Cumulant를 이용한 음성 인식

일반적으로 음성 신호의 전후 데이터들 사이에는 높은 상관관계가 존재하기 때문에 다음 식에 주어진 AR model이 음성 신호 해석에 널리 이용되어 왔다.

$$s(n) + \sum_{i=1}^p a_i \cdot s(n-i) = e(n) \quad (1)$$

잡음 환경에서 사용되는 음성 인식 시스템의 선결 과제는 음성에 포함된 잡음에 대한 면역성이 큰 특징 벡터를 추출해 내는 것이다. 기존의 autocorrelation 방법이나 covariance 방법등과 같이 2차 statistics를 이용하여 AR 계수를 추정한 후 이를 특징 벡터로 하는 방법은 백색 잡음 환경 하에서는 비교적 잘 작동하지만 잡음들 사이에 상관관계가 존재하는 유색 잡음 환경에서는 음성 인식 시스템의 성능이 급격히 떨어진다. 이것은 데이터들 사이의 2차 상관관계로 특징 벡터를 추출하는 기존의 방법들이 음성신호에 포함된 유색 잡음을 음성신호의 일부분으로 간주함으로써 기본적으로 두 신호를 구분해 낼 수 있는 능력이 없기 때문이다. 따라서, 음성신호에 포함된 유색 잡음을 얼마나 제거시킬 수 있느냐가 특징벡터를 구성할때 중요한 요소이며, 3차 이상의 고차 statistics의 몇가지 중요한 성질을 이용할 경우 위의 문제를 효과적으로 해결할 수 있다.

고차 statistics에서 특성 함수(characteristic function)는 많은 정보를 포함하고 있으며 랜덤 변수 X에 대한 moment와 cumulant는 다음의 특성 함수를 이용하여 정의된다.

$$\Phi_x(W) = E[\exp j(WX)] \quad (2)$$

여기서,

$$X = [x_1, x_2, \dots, x_n]^T$$

$$W = [w_1, w_2, \dots, w_n]$$

이때  $k$ 차 moment 및 cumulant를 (2)식의 특성 방정식을 이용하여 정의하면 다음과 같다.

$$Mom[x_1^{k_1}, x_2^{k_2}, \dots, x_n^{k_n}] = (-j)^k \frac{\partial^k \cdot \Phi_x(W)}{\partial w_1^{k_1} \partial w_2^{k_2} \dots \partial w_n^{k_n}} \Big|_{w_1=w_2=\dots=0} \quad (3)$$

$$Cum[x_1^{k_1}, x_2^{k_2}, \dots, x_n^{k_n}] = (-j)^k \frac{\partial^k \cdot \ln[\Phi_x(W)]}{\partial w_1^{k_1} \partial w_2^{k_2} \dots \partial w_n^{k_n}} \Big|_{w_1=w_2=\dots=0} \quad (4)$$

여기서,

$$k = k_1 + k_2 + \dots + k_n$$

일반적으로 3차 미만에서는 평균이 0인 신호의 moment와 cumulant가 동일하며, 그중 특히 2차 moment 또는 cumulant는 autocorrelation으로 불려진다. 3차 이상의 고차 statistics에서 다음의 두가지 성질은 음성 신호 처리에 있어 중요한 의미를 가진다.

먼저 독립적인 두신호  $x(n)$ 과  $w(n)$ 의 합으로 이루어진 신호  $s(n)$ 의 cumulant는 각 신호의 cumulant의 합으로 표시된다는 점이다. 이 점은 2차 statistics를 이용한 autocorrelation 방법이나 covariance 방법으로는 불가능했던 음성 신호와 유색 잡음의 분리를 가능하게 한다. 즉,

$$\begin{aligned} s(n) &= x(n) + w(n) \\ Cum[s_1, s_2, \dots, s_n] &= Cum[x_1 + w_1, x_2 + w_2, \dots, x_n + w_n] \\ &= Cum[x_1, x_2, \dots, x_n] + Cum[w_1, w_2, \dots, w_n] \end{aligned} \quad (5)$$

두번째 중요한 성질은 평균이 0이고 Gaussian 형태의 확률 밀도 함수를 갖는 신호에 대한 3차 이상의 고차 cumulant는 0이라는 점이다.

$$Cum[w_1^{k_1}, w_2^{k_2}, \dots, w_n^{k_n}] = 0, \quad k > 2 \quad (6)$$

일반적으로 음성 데이터는 주기적인 임펄스열이 성도(vocal track)를 통과한 non-Gaussian 특성을 갖는 음성 신호에 평균이 0이고 Gaussian 형태의 확률 밀도 함수를 갖는 잡음이 첨가된 것으로 모델링할 수

있다. 따라서 위의 두가지 중요한 성질을 음성 인식 시스템에 적용하면, 유색 잡음이 첨가된 음성 데이터에서 3차 이상의 고차 cumulant를 이용하여 유색 잡음의 cumulant를 따로 분리해 낼 수 있다. 이때 잡음에 대한 cumulant가 0이 되므로, 결과적으로 잡음에 제거된 음성 신호의 cumulant만이 남게 된다. 즉, non-Gaussian 분포를 갖는 음성 신호를  $x(n)$ , 평균이 0이며 Gaussian 분포를 갖는 잡음을  $w(n)$  이라하면,

$$\begin{aligned} s(n) &= x(n) + w(n) \\ Cum[s_1^{k_1}, s_2^{k_2}, \dots, s_n^{k_n}] &= Cum[x_1^{k_1}, \dots, x_n^{k_n}] + Cum[w_1^{k_1}, \dots, w_n^{k_n}] \end{aligned}$$

여기서,

$$Cum[w_1^{k_1}, \dots, w_n^{k_n}] = 0, \quad k > 2$$

그러므로,

$$Cum[s_1^{k_1}, s_2^{k_2}, \dots, s_n^{k_n}] = Cum[x_1^{k_1}, x_2^{k_2}, \dots, x_n^{k_n}] \quad (7)$$

따라서 이 고차 cumulant  $Cum[s_1^{k_1}, \dots, s_n^{k_n}]$ 를 이용하여 추출한 특징 벡터는 순수한 음성  $x(n)$  만의 정보를 가지고 있기 때문에 이 특징 벡터를 음성 인식 시스템의 입력으로 사용할 경우 유색 잡음 환경에서 일어나는 시스템의 급격한 성능 저하를 막는 효과를 기대할 수 있다.

3차 cumulant를 이용한 특징 벡터의 추출에는 데이터를 프레임 단위로 처리하는 batch-job 형태의 알고리즘으로 TOR(The Third\_order Recursion) 방법과 CTOM<sup>[1]</sup>(The Constrained Third\_order Mean) 방법이 있다. (1)식을 3차 cumulant  $R(k, l) = E[s(n) \cdot s(n+k) \cdot s(n+l)]$ 를 이용하여 다시 쓰면 다음과 같다<sup>[1]</sup>.

$$R(-k, -l) + \sum_{i=1}^p a_i \cdot R(i-k, i-l) = \beta \cdot \delta(k, l) \quad (8)$$

여기서,

$$E[e(n) \cdot e(n+k) \cdot e(n+l)] = \beta \cdot \delta(k, l)$$

$\beta$ : 예측오차  $e(n)$ 의 3승치

$\delta$ : 3차원 임펄스 함수

여기서 예측 오차  $e(n)$ 은 음성 신호  $x(n)$ 에 첨가된 잡음  $w(n)$ 으로 볼 수 있다. 또한 (8)식은  $k=l$ 을 만족하는  $2p+1$  개의 3차 cumulant로 표현할 수 있으며 이것을 벡터 형식으로 다시 쓰면 다음과 같다<sup>[11][3]</sup>.

$$R_c \cdot A = B \tag{9}$$

여기서,

$$R_c = \begin{bmatrix} R(0,0) & R(1,1) & \dots & R(p,p) \\ R(-1,-1) & R(0,0) & \dots & R(p-1,p-1) \\ \vdots & \vdots & \ddots & \vdots \\ R(-p,-p) & R(-p+1,-p+1) & \dots & R(0,0) \end{bmatrix}$$

$$A = [1, a_1, a_2, \dots, a_p]^T$$

$$B = [\beta, 0, 0, \dots, 0]^T$$

따라서 입력 신호  $s(n)$ 에 대한 3차 cumulant  $R(k, l)$ 와 예측 오차  $e(n)$ 의 3승치  $\beta$ 를 알면 벡터  $A$ 를 구할 수 있다. 이 벡터  $A$ 가 음성 인식 시스템의 입력 단계에 쓰일 특징 벡터가 된다. CTOM<sup>[1]</sup> 방법은 변형된 3차 cumulant로 특징 벡터를 구하는 방법이다. 먼저  $q_n(k, i)$ 를 다음과 같이 정의한다.

$$q_n(k, i) = s(n-i) \cdot s^2(n-k), \quad i, k = 1, 2, \dots, p \tag{10}$$

이때 (10)번 식의 평균은 3차 cumulant로 표현할 수 있다.

$$E[q_n(k, i)] = R(i-k, i-k) \tag{11}$$

여기서 잡음의 대한 3차 cumulant  $E[w(n) \cdot w(n-k) \cdot w(n-i)]$ ;  $i, k = 1, 2, \dots, p$ , 는 0이므로 (11)번 식을 3차 cumulant를 이용한 (8)번 식의 AR 예측기에 적용하면 다음과 같이 정의된다.

$$E[q_n(k, 0)] + \sum_{i=1}^p a_i \cdot E[q_n(k, i)] = 0, \tag{12}$$

$$i, k = 1, 2, \dots, p$$

(12)식을 벡터 형식으로 표현하면 다음과 같다.

$$Q \cdot A = q \tag{13}$$

여기서,

$$Q = \begin{bmatrix} q_{11} & \dots & q_{1p} \\ \vdots & \ddots & \vdots \\ q_{p1} & \dots & q_{pp} \end{bmatrix}$$

$$A = [a_1, a_2, \dots, a_p]^T$$

$$q = [q_{10}, q_{20}, \dots, q_{p0}]^T$$

$$q_{ij} = \sum_{n=p+1}^N q_n(i, j)$$

$$q_n(i, j) = s(n-j) \cdot s^2(n-i)$$

이상과 같이 batch-job 형태의 CTOM 알고리즘은 음성 신호의 모음과 같이 일정한 구간 동안 음성 파형이 거의 변하지 않는 시불변 상황에서 기존의 방법과 비교해서 그다지 많이 복잡하거나 계산량을 많이 요구하지 않으면서 비교적 간단하게 구현할 수 있다.

신경 회로망은 인간 뇌신경세포의 정보처리 메커니즘을 모방한 것으로 단순 Perceptron, 다층망 Perceptron, Hopfield 네트워크, 볼츠만 머신 등이 있다<sup>[8][9]</sup>. 이 중에서 주로 패턴 분류를 목적으로 사용되는, 비교적 구현이 용이하고 구조가 간단한, 다층망으로 구성된 Perceptron 모델은 그림 1에서 보는 바와 같이 입력층과 출력층 사이에 중간 (hidden) 층이라 불리는 하나 이상의 계층으로 구성되며 계층간에 연결은 상위를 향하는 방향(feedforward)이다.

음성의 특징벡터를 몇가지 고유한 패턴으로 분리하려고 할 때, 7가지 한국어 단모음은 7가지의 음성 패턴으로 분리가 가능하다. 또한 신경 회로망을 이용한 음성인식은 학습 단계와 인식 단계로 나뉘며, 인식하고자 하는 실제 데이터와 학습 데이터 사이의 일

치하는 정도가 음성 인식 시스템의 성능을 좌우한다. 따라서 자신의 고유한 패턴의 특성을 그대로 간직하고 있는, 잡음에 대한 면역성이 강한 특징벡터를 사용하는 것이 보다 나은 인식율을 보장한다. 앞서 언급한 바와 같이 3차 cumulant 방법은 기존의 방법보다 잡음에 대한 영향을 적게 받기 때문에, 이에 의한 특징벡터 구성이 전체적으로 유색 잡음에 대한 면역성이 강한 인식 시스템을 형성케 한다.

학습 방법은 출력층 오차의 자승치를 최소화 하는 방향으로 중간층과 출력층 사이의 결합 계수를 수정하고, 이어서 입력층과 중간층을 연결하는 결합 계수를 수정하는 역전파(Back Propagation)<sup>[8][9]</sup> 학습 방법이 있다. 본 논문에서는 앞서의 3차 cumulant를 이용한 방법 그리고 기존의 2차 statistics를 이용한 방법으로 추출한 특징 벡터를 이 BP 신경 회로망의 입력으로하여 학습과 인식 실험을 하였으며 이를 통하여 성능을 비교 분석한다.

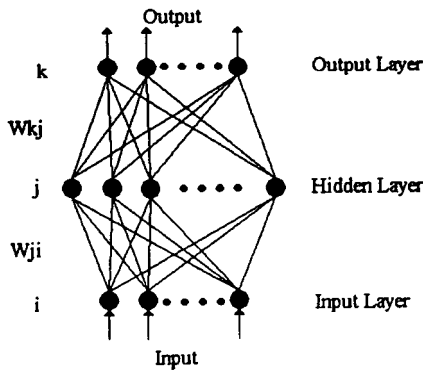


그림 1. 다층망 Perceptron 신경 회로망  
Fig 1. Multi-Perceptron Neural network

### III. 모의실험

본 모의실험에서는 특징벡터를 추출하는 방법으로 3차 cumulant를 이용한 CTOM 방법과 기존의 2차 statistics를 이용한 covariance 방법을 사용하였다. 모의실험에 사용한 데이터는 12차 AR filter로 합성된 음성 신호에 유색 Gaussian 잡음을 첨가하여 얻은 7개의 한국어 합성 단모음 /아/, /에/, /이/, /오/, /우/, /으/, /어/ 각기 50개씩 총 350개이다. 여기서 유색 잡

음  $w(n)$ 은 Gaussian 분포를 갖는 랜덤 신호  $z(n)$ 을 식 (13)의 3차 IIR filter를 통과시켜 발생하였다.

$$w(n) = z(n) + 0.5 \cdot w(n-1) + 0.25 \cdot w(n-2) + 0.125 \cdot w(n-3) \quad (13)$$

첫번째 실험은 먼저 잡음을 증가시키면서 음성 합성 시 사용했던 계수 값과 3차 cumulant 그리고 covariance 방법으로 추출한 계수 값의 편차(deviation)와 bias를 비교하였다. 그림 2는 모음 /아/에 대하여 SNR을 증가시키면서 각각의 방법으로 추정된 AR 계수를 보인다. 그림 2에서 알 수 있듯이 3차 cumulant 방법이 covariance 방법보다 잡음이 커질수록 편차는 크지만 bias가 작은 것을 알 수 있다. 즉, covariance 방법은 잡음이 커져도 편차는 상대적으로 상당히 적어서 안정된 값을 유지하지만 평균값이 원래의 계수 값과는 상당한 차이가 있어 3차 cumulant를 이용하여 추출한 특징벡터가 보다 잡음에 강하다는 것을 알 수 있다.

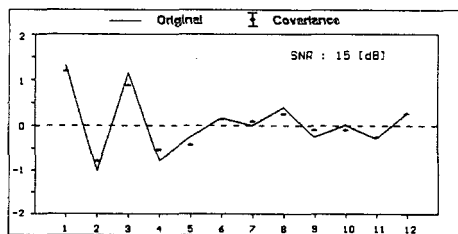
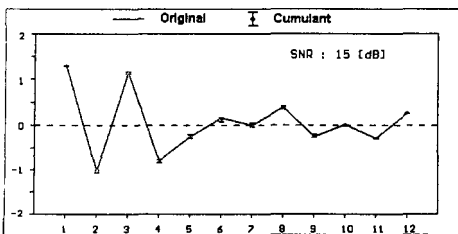
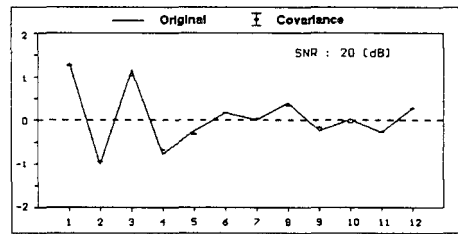
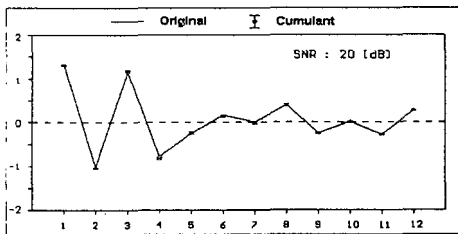
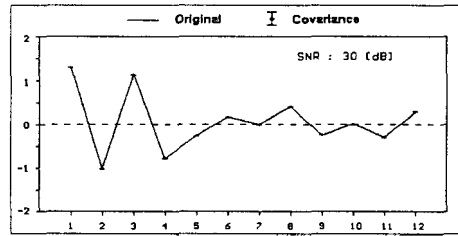
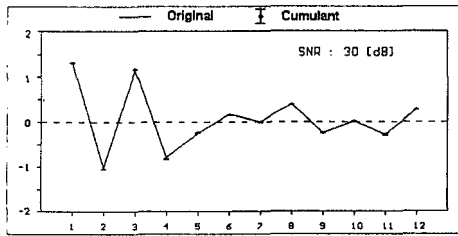
두번째 실험은 데이터 수를 증가시키면서 음성 합성 시 사용했던 계수 값과 각각의 방법으로 추출한 특징벡터를 비교하였다. 그림 3은 이때의 결과를 나타낸 것으로 SNR이 각각 15 dB, 10 dB, 5 dB에서의 데이터 수에 따른 bias를 보이고 있으며 하단 그래프에는 그때의 deviation을 나타낸 것이다. 첫번째 실험에서도 보았듯이 3차 cumulant 방법으로 추출된 특징벡터가 covariance 방법으로 추출된 특징벡터에 비하여 deviation은 크지만 bias가 적은 것을 알 수 있다. 또한 잡음이 증가하면서 covariance 방법으로 추출된 계수 값은 원래 계수 값과의 차이가 급격히 벌어지는 반면 3차 cumulant 방법의 경우 그 계수 값들 사이의 차이가 앞서의 방법에 비하여 비교적 완만하게 벌어져 3차 cumulant 방법이 covariance 방법에 비하여 잡음에 강하다는 것을 알 수 있다. 특히 10 dB와 5 dB 경우 3차 cumulant 방법으로 추출한 특징벡터는 데이터 수가 증가함에 따라 deviation이 점차로 줄어드는 반면 covariance 방법의 경우 데이터 수가 증가함에 따라 오히려 deviation도 증가하여 데이터의 수가 30,000개를 전후해서는 각각의 방법으로 인한 차이가 거의 없음을 볼 수 있다. 이것은 기본적으로 유색 잡음을 제거할 수 있는 능력이 없는 covariance 방법의 경우 유색 잡음이 일정 정도 이상

커지면 데이터 수가 증가함에 따라 잡음에 의한 영향이 계속해서 누적되기 때문이다. 반면에 3차 cumulant 방법의 경우 데이터 수가 증가함에 따라 잡음에 의한 영향보다는 원래의 음성신호에 의한 영향이 증가함으로써 deviation이 조금씩 줄어들음을 볼 수 있다. 따라서 3차 cumulant 방법으로 추출한 특징벡터는 특히 잡음이 크기가 비교적 크고 잡음들 사이에 상관관계가 높은 상황에서 보다 효과적임을 알 수 있다.

세번째 실험에서는 입력층과 출력층 사이에 1개의 은닉층을 갖는 다층구조 신경망을 이용하여 인식률을 실험하였다. 입력층의 노드는 13개로 하였으며 중간층 노드 9개, 출력층 노드는 7개로 하였다. 신경망의 학습을 위하여 BP 학습 알고리즘을 사용하였으며 여기에 local minimum에 빠지는 것을 방지하기 위해 수정된 모멘트 법과 학습 속도를 개선시킬 수 있는 일괄 수정법을 적용하였다. 인식 실험시 출력층의 크기가 가장 큰 것을 인식한 것으로 하였다. 시그모이

드 함수의 기울기는 1.6,  $\alpha$ ,  $\beta$ 는 모두 1.0으로 하였다.

학습 데이터는 음성 합성시 사용했던 특징벡터를 사용하였으며 인식률은 합성음에 잡음의 크기를 변화하면서 각각의 방법으로 추정된 특징 벡터를 이용하여 계산하였다. 표 1은 이때의 결과를 나타낸 것이다. 표 1에서 알 수 있듯이 covariance 방법의 경우 15 dB을 전후로 해서 인식율이 급격히 떨어짐으로써 잡음에 대한 면역성이 크게 떨어진다는 것을 알 수 있다. 반면에 cumulant를 이용하여 추출한 특징 벡터를 입력으로 하는 음성 인식 시스템의 경우 급격한 인식률의 저하는 나타나지 않으며 모든 모음에 대해서 비교적 고른 인식율을 보여준다. 따라서 3차 cumulant를 이용한 방법으로 추출한 특징 벡터가 SNR이 15 dB을 전후로 해서 covariance 방법으로 추출한 특징 벡터에 비하여 잡음에 대한 면역성이 크다는 것을 알 수 있다.



(a)

(b)

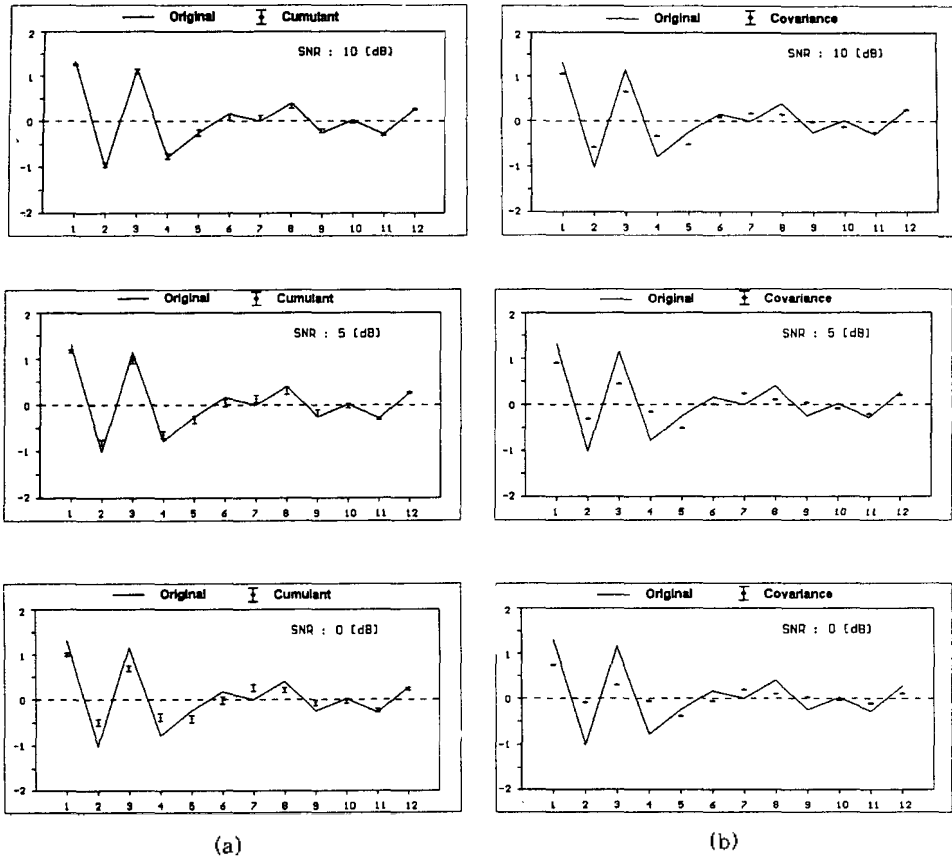
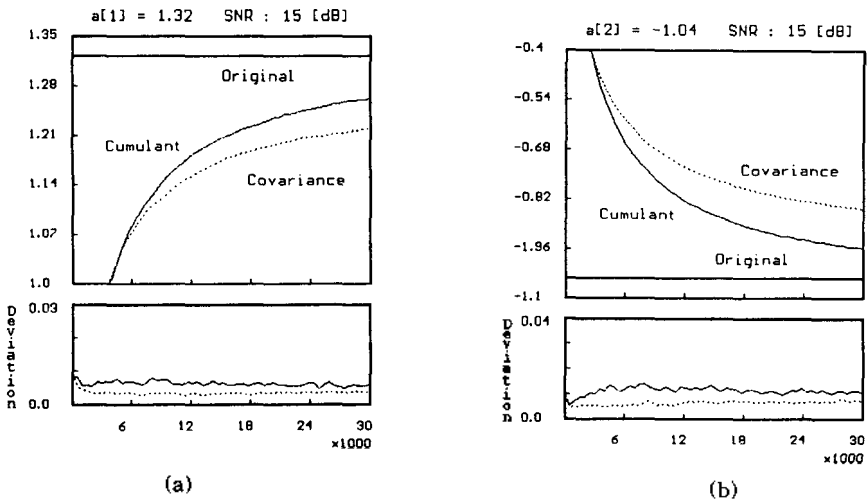


그림 2 각 SNR에서 추정된 AR 파라미터의 편차 및 bias (음성 신호 /아/의 경우)

(a) cumulant 방법 (b) covariance 방법

Fig 2. Deviation and bias of the estimated AR coefficients at various SNR (in case of /a/)

(a) cumulant method (b) covariance method



(a)

(b)

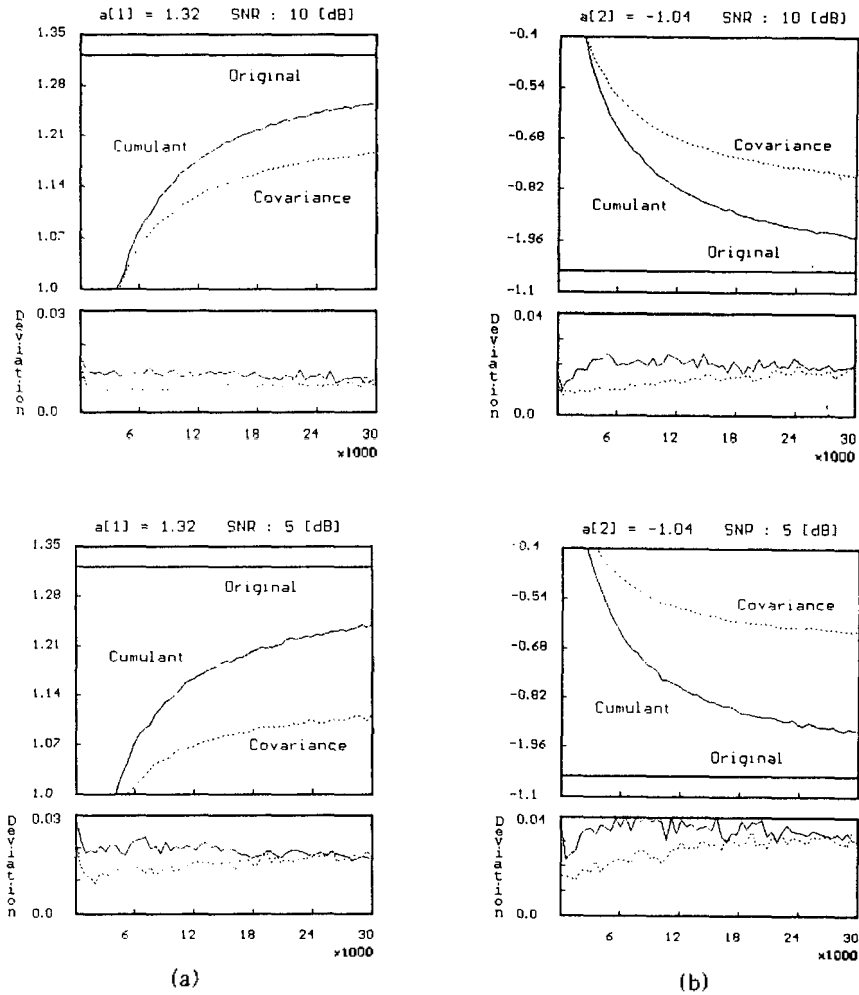


그림 3. 각 SNR에서 데이터 갯수에 따른 AR 파라미터 비교 (음성 신호 /아/의 경우)  
 (a) 첫번째 계수 (b) 두번째 계수

Fig 3. AR coefficients vs. number of data at various SNR (in case of /a/)  
 (a) first coefficient (b) second coefficient

표 1. SNR에 따른 인식률 비교

Table 1. Recognition rate vs. SNR

30 [dB]	CUMULANT / COVARIANCE							인식률(%)
	/아/	/에/	/이/	/오/	/우/	/으/	/어/	
/아/	50/50							100.0/100.0
/에/		50/50						100.0/100.0
/이/			50/50					100.0/100.0
/오/				50/50				100.0/100.0
/우/					50/50			100.0/100.0
/으/						50/50		100.0/100.0
/어/							50/50	100.0/100.0
총 계								100.0/100.0

(a) 30dB

20 [dB]	CUMULANT / COVARIANCE							인식률(%)
	/아/	/에/	/이/	/오/	/우/	/으/	/어/	
/아/	49/49						1/1	98.0/98.0
/에/		50/50						100.0/100.0
/이/			50/50					100.0/100.0
/오/				50/50				100.0/100.0
/우/					50/50			100.0/100.0
/으/						50/50		100.0/100.0
/어/							50/50	100.0/100.0
총 계								99.7/99.7

(b) 20dB



IV. 결 론

음성 인식 시스템의 실용화 문제에서 반드시 필요한 것이 잡음에 대한 면역성이며 그중에서도 특히 유색 잡음에 대한 연구가 반드시 필요하다. 모의 실험을 통하여 유색 잡음 환경에서 3차 cumulant를 이용하여 구한 특징 벡터를 입력으로 하는 음성 인식 시스템의 인식율은 기존의 2차 statistics를 이용한 방법보다 높은 인식율을 나타냄으로써 잡음 환경에서의 급격한 성능 저하를 막을 수 있는 가능성을 보여주었다. 본 논문은 고차 statistics를 이용한 단모음 인식에 국한되어 있으며 이를 이용한 연속단어 인식에 대한 연구는 현재 진행중이다.

참 고 문 헌

1. C.L. Nikias and M.R. Raghuveer, "Bispectrum Estimation: A Digital Processing Framework," *Proceeding of the IEEE*, Vol.75, No.7, pp.869-891, July, 1987.
2. C.L. Nikias, *Higher-Order Spectra Analysis*, Prentice-hall International Inc, 1993.
3. C.L. Nikias and M.R. Raghuveer, "Parametric Approach to Bispectrum Estimation," *ICASSP*, 38.1. 1-38.1.4, 1984.
4. C.L. Nikias and M.R. Raghuveer, "Bispectrum Estimation via AR Modeling," *Signal Processing 10*, pp. 35-48, 1986.
5. R. Fulchiero and A.S. Spanias, "Speech Enhancement Using the Bispectrum," *ICASSP*, Vol.4, pp. 488-491, 1993.
6. E. Masgrau, J.A. Rodrigues-Fonollosa, and A. Ardany, "Enhancement of Speech by Using Higher Order Spectral Modeling," *Signal Processing VI*, pp. 307-310, 1992.
7. A.N. Delopoulos and G.B. Giannakis, "Strongly Consistent Identification Algorithms and Noise Insensitive MSE Criteria," *Proceeding of the IEEE*, Vol. 40, No.8, pp.1955-1922, August, 1992.
8. J.M.Zurada, *Introduction to Artificial Neural Systems*, West Publishing Company, 1992.
9. J.A. Freeman and D.M. Skapura, *Neural Networks Algorithms, Applications, Programming Techniques*, Addison-Wesley Publishing Company, 1991.

15 [dB]	CUMULANT / COVARIANCE							인식률(%)
	/아/	/에/	/이/	/오/	/우/	/으/	/어/	
/아/	49/49						1/1	98.0/98.0
/에/		50/48				/2		100.0/96.0
/이/		6/	44/50					88.0/100.0
/오/				50/			/50	100.0/0.0
/우/					50/50			100.0/100.0
/으/						50/50		100.0/100.0
/어/							50/50	100.0/100.0
총 계								98.0/84.6

(c)15dB

10 [dB]	CUMULANT / COVARIANCE							인식률(%)
	/아/	/에/	/이/	/오/	/우/	/으/	/어/	
/아/	49/49						1/1	98.0/98.0
/에/		44/				6/50		88.0/0.0
/이/		24/	23/50	3/				46.0/100.0
/오/				47/			3/50	94.0/0.0
/우/					50/50			100.0/100.0
/으/						50/50		100.0/100.0
/어/							50/50	100.0/100.0
총 계								89.4/71.1

(d)10dB

5 [dB]	CUMULANT / COVARIANCE							인식률(%)
	/아/	/에/	/이/	/오/	/우/	/으/	/어/	
/아/	49/						1/50	98.0/0.0
/에/		20/				30/50		40.0/0.0
/이/		6/	14/		30/50			28.0/0.0
/오/				12/		/48	38/2	24.0/0.0
/우/					50/	/50		100.0/0.0
/으/						50/50		100.0/100.0
/어/							/50	100.0/100.0
총 계								70.0/14.286

(e)5dB

0 [dB]	CUMULANT / COVARIANCE							인식률(%)
	/아/	/에/	/이/	/오/	/우/	/으/	/어/	
/아/	46/						4/50	92.0/0.0
/에/	3/	13/	4/			30/50		26.0/0.0
/이/			1/		49/50			2.0/0.0
/오/							50/50	0.0/0.0
/우/					49/	1/50		98.0/0.0
/으/						50/50		100.0/100.0
/어/							/50	100.0/100.0
총 계								59.7/14.3

(f)0dB

▲이 형 근(Hyung-Gun Lee) 1966년 11월 10일생



1985년 3월~1991년 2월: 중앙  
대학교 공과대학 전  
기공학과 전기공학  
학사  
1992년 9월~1994년 4월: 중앙  
대학교 공과대학 전  
기공학과 대학원 석  
사과정

▲조 용 수

현재: 중앙대학교 전자공학과 조교수  
(1993년 12권 1E호 참조)

▲양 권 영(Won-Young Yang) 1953년 10월 6일생



1971년 3월~1975년 2월: 서울  
대학교 공과대학 전  
기공학과 전기공학  
학사  
1981년 9월~1986년 8월: 미국  
Southern Califor-  
nia 대학 전기공학과  
전기공학 박사

1986년 9월~1994년 4월: 중앙대학교 전기공학과 교수