

# 신호편의제거 알고리즘에 기초한 강인한 음성 인식시스템의 구현

## On Implementing a Robust Speech Recognition System Based on a Signal Bias Removal Algorithm

임 계 종\*, 계 영 철\*, 구 명 완\*\*

( Kye Jong Yim\*, Young Chul Kay\*, Myoung Wan Koo\*\*)

\* 이 연구는 97년도 한국학술진흥재단 연구비 지원에 의한 결과임(997-001-E0029).

### 요 약

본 논문에서는 오염된 입력음성을 보상하는 방법들 중 신호편의제거 알고리즘을 근간으로 하여 환경에 독립적이고, 연산량을 최소화하며, 그리고 기존의 인식시스템에 쉽게 적용시킬 수 있는 새로운 보상 알고리즘을 개발하였다. 이를 위하여 기존의 신호편의제거 알고리즘에 대하여 여러 개의 바이어스를 동시에 사용하는 알고리즘과 코드북을 부분적으로 검색하는 알고리즘을 추가로 제안하였다. 실험 결과, 본 논문에서 제안한 두 가지 알고리즘을 복합하여 적용한 경우, 기존 알고리즘에 비하여 약 1/8정도로 연산량을 감소시킬 수 있었으며, 보상 후의 인식률도 기존 신호편의제거 알고리즘의 77.58%에서 81.32%로 향상되었다.

### ABSTRACT

Particularly based on the signal bias removal(SBR) algorithm for compensating the corrupted speech, this paper presents a new algorithm which is independent of environments, minimizes the amount of computation, and is readily applicable to the conventional recognition system. To this end, a multiple-bias algorithm and a partial codebook search algorithm have been added to the conventional SBR algorithm. The simulation results show that combining the two algorithms proposed in this paper provides a reduction of computation time to 1/8 times as well as an improvement of the recognition rate from 77.58% of the conventional system to 81.32%.

### I. 서 론

일반적인 음성인식 시스템의 경우 SNR이 25dB 이상인 환경에서는 인식률이 거의 일정하게 나타난다. 그러나, 잡음이 그 이상 증가하게 되면 급격한 인식률의 저하를 보이게 된다[4]. 특히, 음성인식 시스템의 특징벡터를 구하기 위해 널리 사용되는 LP분석의 경우 잡음에 아주 민감한 특성을 나타낸다. 잡음환경에서 음성인식 시스템의 성능을 향상시키기 위하여 지금까지 많은 연구가 이루어져 왔지만, 실제 잡음환경을 적절히 모델링하는 것이 매우 어렵기 때문에 대부분의 시스템이나 알고리즘들은 작은 크기의 잡음이나 stationary한 백색잡음을 가정하고 있다. 따라서, 실험실과 실제 환경 사이에서의 인식시스템의 성능

에 큰 차이를 종종 보여 왔다.

지금까지의 많은 보상 알고리즘들은 환경에 독립적이지 못하여 새로운 환경에서는 적용이 불가능하거나, 보상 성능이 크게 저하되는 문제점을 가지고 있다. 또한 이러한 보상 알고리즘들은 상당한 계산량을 요구하기 때문에 기존 인식 시스템에 적용하기 힘든 문제점을 지니고 있다. 이러한 문제점들을 해결하기 위하여 본 논문에서는 미지의 새로운 환경에서도 기존 인식시스템의 성능을 최대한 유지할 수 있는, 환경에 독립적인 보상 알고리즘을 제안하고자 한다. 본 논문은 특별히 기존의 신호편의제거 알고리즘을 기반으로 하여 앞서 언급한 문제점들을 해결하고자 한다.

본 논문에서는 기존 신호편의제거 알고리즘의 성능을 좀더 개선하기 위하여 다음과 같은 2가지 방법을 수정제안 한다: 하나의 바이어스 대신에 입력 프레임의 SNR에 따라 서로 다른 바이어스를 사용하는 방법을 제안한다; 바이어스 추정을 위하여 전체 코드북을 검색하는 기존의

\* 홍익대학교 전자공학과

\*\* 한국통신

접수일자: 1999년 9월 11일

벡터양자화 알고리즘을 수정하여, 코드워드간의 거리에 따라 제한된 수의 코드워드만을 이용하여 벡터양자화를 수행하는 새로운 벡터양자화 알고리즘을 제안한다. 기존 신호편의제거 알고리즘의 경우 벡터양자화가 대부분의 연산량을 차지하므로, 제안된 새로운 벡터양자화 알고리즘을 이용할 경우 연산량을 크게 줄일 수 있다.

## II. 본 론

### 2.1. 신호편의제거 알고리즘

신호편의제거 알고리즘은 처음 Cox와 Bridle에 의하여 화자적응의 목적으로 사용되었으며[1], Rahim과 Juang은 이것을 전화망의 채널특성을 보상하기 위하여 사용하였다[2].

전화망에 의한 채널왜곡, 수화기 마이크의 특성에 의한 왜곡 등 주변환경에 의한 왜곡을  $H(\omega)$ 이라 하고, 주파수 영역에서 부가적인 잡음을  $N(\omega)$ 으로 볼 때, 음성 인식시스템에 입력되는 오염된 음성은 다음과 같이 표현할 수 있다.

$$Y(\omega) = H(\omega) \cdot X(\omega) + N(\omega) \quad (1)$$

식 (1)에서  $X(\omega)$ 는 오염되지 않은 음성신호를 나타내며  $Y(\omega)$ 는 오염된 입력음성을 나타낸다. 사용자가 인식시스템에 접속된 상태에서는  $H(\omega)$ 와  $N(\omega)$ 이 일정하다고 가정할 수 있으며, 이를 입력음성의 바이어스(bias, 편이)라고 볼 수 있다. 신호편의제거 알고리즘은 오염된 입력 음성 신호로부터  $H(\omega)$ 를 바이어스로 추정하여 제거함으로써 오염된 음성신호를 보상하여 준다. 오염된 음성의 바이어스를 추정하기 위한 방법으로는 ML(Maximum Likelihood) 추정방법을 사용한다.

다음과 같이 주어진 음성인식 시스템의 모델  $\Lambda$ 에서 입력 음성의 특징벡터열  $X$ 가 관측될 확률밀도함수(probability density function)로 likelihood를 정의할 수 있다.

$$p(X|\Lambda) = \prod_i \max_{\lambda_i} p(x_i|\lambda_i) \quad (2)$$

여기서  $\Lambda$ 는 음성인식 시스템의 모델을 나타내며,  $\lambda_i$ ,  $i=1,2,\dots,M$ 는 모델  $\Lambda$ 의 부모델(sub-model)들이다.  $X = x_1, x_2, \dots, x_b, \dots, x_T$ 는 오염되지 않은 음성의 특징벡터열을 나타내며,  $t$ 는 프레임 인덱스(frame index)이다. 식 (2)에서  $\lambda_i$ 가 동일한 확률을 갖는 경우 식 (2)의 likelihood 함수를 벡터양자화와 동일하게 생각할 수 있으며, 이때  $\lambda_i$ 는 벡터양자화 코드북의 코드워드에 해당하게 된다.

식 (1)에서  $N(\omega)$ 를 충분히 최소화 시킨 후, 식 (1)을 cepstrum으로 표현하면,

$$Y = X + b \quad (3)$$

여기서  $Y$ 는  $y_1, y_2, \dots, y_t, \dots, y_T$ 로서 오염된 음성의 cepstrum 특징벡터열을, 그리고  $b$ 는  $H(\omega)$ 의 cepstrum을 나타낸다. 현재 추정된 바이어스를  $b$ 라고 가정하면

$$p(Y|b) = p(Y - b) \quad (4)$$

따라서, 식 (2)의 likelihood 함수는 다음과 같이 나타낼 수 있다.

$$p(Y|b, \Lambda) = \prod_i \max_{\lambda_i} p(y_i - b | \lambda_i) \quad (5)$$

$$p(Y|\bar{b}, \Lambda) = \max_b p(Y|b, \Lambda) \quad (6)$$

여기서  $\Lambda$ 는 전체 코드북을 나타내며,  $\lambda_i$ ,  $i=1,2,\dots,M$ 는 코드북  $\Lambda$ 를 구성하는 cluster들이다. 식 (5)는 오염된 입력음성을 현재 바이어스로 보상을 취한 신호에 대한 maximum likelihood를 의미한다. 식 (5)에 EM(Estimation-Maximization) 알고리즘을 적용하여, 식 (6)과 같이 현재 보상된 신호의 벡터양자화 에러를 최소화하는 새롭게 추정된 바이어스  $\bar{b}$ 를 구할 수 있다.

신호편의제거 알고리즘에서 사용하는 EM 알고리즘은 다음과 같다: 현재 추정된 바이어스 값을  $b$ 라고 하면,  $b$ 를 이용하여 보상된 신호  $\hat{x}_i$ 는 식 (7)과 같이 나타낼 수 있다.

$$\hat{x}_i = y_i - b, \quad 1 \leq i \leq T \quad (7)$$

식 (7)의 보상된 신호  $\hat{x}_i$ 에 대한 특징벡터 양자화기의 가장 근접한 이웃(nearest neighbor), 즉 벡터양자화의 결과는,

$$i^* = \arg \max_i p(y_i | b, \lambda_i) = \arg \max_i p(\hat{x}_i | \lambda_i) \quad (8)$$

$$z_i = \mu_{i^*}, \quad 1 \leq i \leq T \quad (9)$$

식 (8)에서  $i^*$ 는 코드북 인덱스를 나타내며,  $\mu_{i^*}$ 는 벡터양자화 코드북의 코드워드, 즉 cluster  $\lambda_{i^*}$ 의 중심(centroid)에 해당한다.  $\hat{x}_i$ 에 대한 벡터양자화의 결과  $z_i$ 를 이용한 바이어스 추정은 식 (10)과 같다.

$$\bar{b} = \frac{1}{T} \sum_{i=1}^T (y_i - z_i) \quad (10)$$

식 (7),(8),(9),(10)에 의해 추정된 바이어스가 부분최적해(local optimal solution)가 될 때까지 바이어스 추정을 반복하여 오염된 입력음성 보상에 필요한 바이어스를 구

할 수 있다 [2].

2.2. 제안한 알고리즘

2.2.1 N개의 바이어스를 이용하는 신호편의제거 알고리즘

SDCN (SNR-Dependent Cepstral Normalization) 알고리즘에서는 오염된 음성의 켈스트럼  $y$ 로부터 보상된 음성의 켈스트럼  $\hat{x}$ 를 얻기 위하여 다음과 같은 식을 가정한다[3].

$$\hat{x} = y - w \text{ (SNR)} \quad (11)$$

즉, 보상을 위한 벡터  $w$ 는 입력 프레임의 순시신호대 잡음비(instantaneous SNR)에 종속적이라고 가정한다. 여기서 신호대잡음비를  $y[0] - n[0]$ 과 같이 근사화 할 수 있으며, 이 때,  $n[0]$ 는 입력음성의 잡음구간에 해당하는 특징벡터의 평균이 된다. 이때 입력음성의 전체구간에서 잡음의 크기가 일정하다고 가정하면, 입력 프레임의 SNR은 입력프레임의 로그에너지에 비례한다고 볼 수 있다. 즉,

$$\text{SNR} \propto y[0] \quad (12)$$

따라서, 본 절에서 제안한 알고리즘에서는 실험에 의해 결정된 기준 에너지레벨에 대한 입력 프레임의 로그에너지에 따라  $N$ 개의 바이어스를 가정한다.

입력 프레임의 SNR에 따라 현재 추정된  $N$ 개의 바이어스 값을  $b[\text{SNR}_i]$ ,  $i=0, 1, \dots, N$ 라고 하면,  $b[\text{SNR}_i]$ 를 이용하여 보상된 신호  $\hat{x}_i$ 는 식 (13)과 같이 나타낼 수 있다.

$$\hat{x}_i = y_i - b[Q\{\text{SNR}_{y_i}\}], \quad (13)$$

$$Q\{\text{SNR}_{y_i}\} \in \{\text{SNR}_1, \text{SNR}_2, \dots, \text{SNR}_N\}$$

여기서  $Q\{\text{SNR}_{y_i}\}$ 는 입력 음성의  $i$ 번째 프레임  $y_i$ 의 로그에너지를 양자화한 결과를 나타낸다. 즉,  $b[Q\{\text{SNR}_{y_i}\}]$ 는  $\text{SNR}_{y_i}$ 에 해당하는 바이어스를 나타내며, 바이어스의 추정은 다음 식 (14)와 같다.

$$\bar{b}[\text{SNR}_i] = \frac{\sum_{t=1}^T (y_t - z_t) \delta[Q\{\text{SNR}_{y_t}\} - \text{SNR}_i]}{\sum_{t=1}^T \delta[Q\{\text{SNR}_{y_t}\} - \text{SNR}_i]} \quad (14),$$

$$i=0, 1, \dots, N$$

위의 식 (14)에서  $\delta[]$ 는 Kronecker delta이며,  $\delta[Q\{\text{SNR}_{y_t}\} - \text{SNR}_i]$ 은 입력 음성의  $i$ 번째 프레임  $y_t$ 의  $\text{SNR}_{y_t}$ 의 양자화된 값이 현재 추정하고자 하는 바이어스  $b[\text{SNR}_i]$ 의 SNR에 해당하면 1의 값을, 그 외의 경우에

는 0의 값을 갖게 된다. 즉, 현재 추정하고자 하는 바이어스는 그에 해당하는 SNR을 갖는 프레임들에 대하여서만 기존의 방식에 의한 벡터양자화 에러를 평균하여 구해진다.

2.2.2 제한된 수의 코드워드에 대한 벡터양자화를 이용한 신호편의제거 알고리즘

신호편의제거 알고리즘에서 바이어스 추정을 위한 연산의 대부분은 벡터양자화가 차지하며,  $T$ 프레임의 특징벡터를 사용하여  $k$ 번의 바이어스추정을 할 경우  $T \cdot (k+1)$ 번의 벡터양자화가 필요하다. 따라서 보상알고리즘의 연산 시간을 최소화하기 위하여서는 전체 코드북을 검색하는 기존의 벡터양자화 알고리즘을 수정하여 연산량을 줄일 수 있는 새로운 벡터양자화 알고리즘이 필요하다.

신호편의제거 알고리즘에서 벡터양자화 과정은 현재의 바이어스를 이용하여 새로운 바이어스를 추정하기 위해 사용되며 그 과정은 다음과 같이 정리할 수 있다.

오염된 신호  $y_t$ 를 보상하기 위한 바이어스를 추정하기 위해 사용하는 입력 음성의 전체 프레임 수를  $T$ 라고 하고,  $n$ 번째 추정에서 현재 추정된 바이어스를  $\bar{b}^n$ 이라고 하면, 식 (15)와 같이  $\bar{b}^n$ 를 이용하여 오염된 입력  $y_t$ 를 보상한  $\hat{x}_t^n$ 을 계산하고, 이에 대한 벡터양자화를 식 (16)와 같이 수행하여 코드워드  $\mu_{i^*}$ 를 결정한다.

$$\hat{x}_t^n = \hat{x}_t^{n-1} - \bar{b}^n, \quad 1 \leq t \leq T \quad (15)$$

$$i^* = \arg \min_i \| \hat{x}_t^n - \mu_i \| \quad (16)$$

$$\hat{z}_t = \mu_{i^*}, \quad 1 \leq t \leq T$$

여기서, 벡터양자화 결과인  $\hat{z}_t$ 를 이용하여 식 (17)과 같이 새로운 바이어스를 추정한다.

$$\bar{b}^{n+1} = \frac{1}{T} \sum_{t=1}^T (\hat{x}_t^n - \hat{z}_t) \quad (17)$$

위 식을 정리하여 보면,  $n+1$ 번째 추정된 바이어스에 의해 보상된 신호  $\hat{x}_t^{n+1}$ 를 다음과 같이 나타낼 수 있다.

$$\hat{x}_t^{n+1} = y_t - \bar{b}^0 - \bar{b}^1 \dots - \bar{b}^{n+1} \quad (18)$$

$$= \hat{x}_t^n - \bar{b}^{n+1}$$

식 (18)에서 새로 보상된 신호는 이전의 바이어스로 보상된 신호와  $\bar{b}^{n+1}$ 만큼의 차이가 나며, 이때  $\bar{b}^{n+1}$ 는  $\bar{b}^n$ 로 보상된 신호의 벡터양자화 에러를 평균한 것이므로 크기가 크지 않다. 그러므로,  $\hat{x}_t^{n+1}$ 와  $\hat{x}_t^n$ 과의 거리차가 제한되어 있음을 알 수 있다. 따라서, 새로

운 바이어스를 추정하기 위하여 벡터양자화를 하는 경우 이전 바이어스에 의한 벡터양자화 결과를 이용하여 전체 코드북의 일부 코드워드들에 대하여서만 벡터양자화를 수행하여 벡터양자화 연산량을 줄일 수 있다. 즉, 보상 알고리즘을 사용하기 전에 코드북의 각 코드워드들 간의 거리를 미리 계산할 수 있으며, 이 거리를 이용하여 이전 벡터양자화 결과인 코드워드와 가장 가까운 거리에 있는  $L$ 개의 코드워드들만으로 벡터양자화를 제한 할 수 있을 것이다.

이러한 벡터양자화 방법을 다음과 같이 식으로 정리 할 수 있다.

$$i^{n+1} = \arg \min_i \| \hat{x}_i^n - \mu_i \|^2, \quad (19)$$

$$i = \{ \arg \min_j L \| \mu_{j^n} - \mu_j \|^2 \}$$

$$\hat{x}_i = \mu_{j^{n+1}}, \quad 1 \leq i \leq T \quad (20)$$

위의 식 (19)에서  $\mu_{j^n}$ 은  $n$ 번째 바이어스 추정에서의 벡터양자화의 결과를 나타낸다.  $L$ 은 제한된 코드워드의 개수이며,  $\min L$  은 이전 벡터양자화된 코드워드와의 거리가 작은 순서로  $L$ 개의 코드워드들에 대하여 벡터양자화를 수행한다는 것을 의미한다.

### III. 실험 및 결과

#### 3.1. 음성 데이터베이스

신호편의제거 알고리즘을 이용한 인식률측정 실험에는 실험실에서 전화를 이용하여 녹음한 음성 데이터베이스(DB1), 인식시스템을 이용하여 녹음된 음성 데이터베이스(DB2), 그리고 실제 환경에서 인식시스템을 운영하며 녹음된 음성 데이터베이스(DB3)의 세 가지가 사용되었다. 각 데이터베이스는 전화 다이얼링 예약어로 사용될 수 있는 150단어로 구성되며, 서울 및 경기지역에 거주하는 10대에서 50대까지의 남녀의 음성으로 녹음되었다. 인식 시스템의 훈련에는 세 가지 음성 데이터베이스가 모두 사용되었고, 인식률 측정에는 DB3의 음성을 사용하였다.

DB1의 음성은 가정, 사무실, 거리등 여러 장소에서 실제 상황에 가깝도록 녹음하였다. 공중전화망을 통하여 테이프에 녹음된 입력음성은 워크스테이션에서 A/D변환되어 저장되었으며, 327명분의 음성이 인식시스템의 훈련에 사용되었다. DB2는 화자가 인식시스템을 사용하는 과정에서 녹음되었으며, 테이프를 거치지 않고 바로 A/D변환되어 저장되었다. 296명의 음성이 녹음되어, 267명의 음성이 훈련에 사용되었다. 실제 사용환경의 음성 데이터베이스(DB3)는 인식 시스템을 시험 운영하여, 시험 접속자가 인식시스템을 실제 사용하는 과정에서 녹음되었다. 인식 시스템의 훈련에 사용된 DB3의 음성데이터는 화자에 관계없이 모두 7900단어가 사용되었으며, 인식률 측정에는 669단어가 사용되었다.

#### 3.2. 기본 인식시스템

인식시스템의 특징벡터로는 기중 LPC 켈스트랄계수 외에 이것의 일차빼기, 이차빼기, 로그파워의 일차, 이차 빼기 값이 사용되었다. 특징 벡터들을 각각의 코드북에 의해 벡터양자화 되어 사용된다. 이때 사용된 코드북들은 LBG(Linde-Buzo-Gray)알고리즘에 의하여 작성되었으며, 이들 중 3개의 코드북은 256개의 코드워드를 갖고 로그 파워의 벡터양자화에 사용되는 코드북은 64개의 코드워드를 갖는다.

#### 3.3. 실험 결과

##### 3.3.1 N개의 바이어스를 이용하는 신호편의제거 알고리즘

본 논문의 실험에서는 2개의 바이어스를 이용하였으며, 입력 프레임의 SNR에 따라 2개의 다른 바이어스를 이용하게 된다. 실험에서는 입력음성의 잡음성분이 일정하다는 가정하에 입력 프레임의 로그에너지에 따라 바이어스를 구성하였다.

그림 1은 바이어스 추정의 반복회수에 따른 인식률을 측정한 결과이다. 여기서 baseline이란 보상을 안한 인식 시스템을 의미한다. 그래프에서 볼 수 있듯이 바이어스의 추정이 약 6회 이상일 때 기존의 신호편의제거 알고리즘보다 성능이 향상되고 있음을 알 수 있다.

바이어스를 2개 이상 사용하는 방법도 실험하여 보았으나, 바이어스 수를 증가시킴에 따라 각 바이어스를 추정하기 위한 충분한 수의 특징벡터가 확보되지 않아 오히려 성능이 크게 저하되었다. 입력음성이 충분히 긴 경우에는 2개 이상의 바이어스를 사용하는 방법도 고려할 수 있을 것이다.

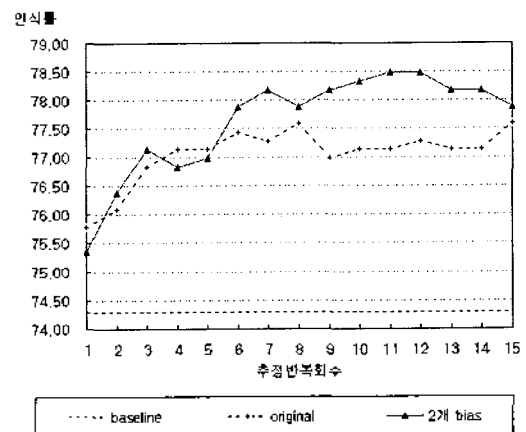


그림 1. 2개의 바이어스를 이용하는 신호편의제거 알고리즘의 인식률

Fig. 1. The recognition rate of the SBR algorithm with 2 biases.

##### 3.3.2 제한된 수의 코드워드에 대한 벡터양자화를 이용한 신호편의제거 알고리즘

제한된 수의 코드워드에 의한 벡터양자화를 이용한 신호편의제거 알고리즘에서는 기존 코드북의 각 코드워드들에 대하여 N개의 이웃한 코드워드들을 미리 계산하여 두며, 바이어스 추정을 위하여 이들 이웃한 코드워드들에 대하여서만 벡터양자화를 수행한다. 표 1은 입력음성에 대하여 바이어스가 변화함에 따라 벡터양자화 결과가 변화하는 과정을 비교한 것이며, 여기서는 임의의 5개 프레임에 대하여 각각 기존의 신호편의제거 알고리즘을 적용한 경우 그리고 16개, 32개의 이웃한 코드워드들에 대하여만 벡터양자화를 수행한 경우를 각각 보여주고 있다. 밑줄이 그어진 부분은 바이어스추정에 따른 벡터양자화의 결과가 수렴하였음을 나타내며, 각 열은 각 프레임을 나타낸다.

표 1. 바이어스 추정에 따른 벡터양자화 결과의 변화  
Table 1. The changes in the results of vector quantizations associated with the bias estimates.

| 추정회수 | 기존 SBR (256 인접코드)          |
|------|----------------------------|
| 1    | 213   233   132   73   212 |
| 2    | 233   233   132   69   212 |
| 3    | 81   233   132   69   212  |
| 4    | 81   233   24   69   212   |
| 5    | 81   97   24   84   212    |
| 6    | 81   97   24   67   192    |
| 7    | 81   97   69   67   192    |
| 8    | 81   97   69   67   192    |
| 9    | 81   97   69   67   192    |
| 10   | 81   97   69   67   192    |
| 11   | 81   97   69   67   192    |
| 12   | 81   97   69   67   192    |
| 13   | 81   97   69   67   192    |
| 14   | 81   97   69   67   192    |
| 15   | 81   97   69   67   192    |

| 추정회수 | 16 인접코드                    | 32 인접코드                    |
|------|----------------------------|----------------------------|
| 1    | 213   233   132   73   212 | 213   233   132   73   212 |
| 2    | 212   233   132   69   212 | 212   233   132   69   212 |
| 3    | 133   233   132   69   212 | 101   237   132   69   212 |
| 4    | 233   233   24   69   212  | 81   97   24   69   212    |
| 5    | 101   233   24   69   212  | 81   97   24   84   212    |
| 6    | 81   97   24   69   212    | 81   97   24   67   192    |
| 7    | 81   97   24   84   212    | 81   97   69   67   192    |
| 8    | 81   97   24   84   212    | 81   97   69   67   192    |
| 9    | 81   97   24   84   192    | 81   97   69   67   192    |
| 10   | 81   97   24   84   192    | 81   97   69   67   192    |
| 11   | 81   97   69   84   192    | 81   97   69   67   192    |
| 12   | 81   97   69   82   192    | 81   97   69   67   192    |
| 13   | 81   97   69   67   192    | 81   97   69   67   192    |
| 14   | 81   97   69   67   192    | 81   97   69   67   192    |
| 15   | 81   97   69   67   192    | 81   97   69   67   192    |

표 1에서 볼 수 있듯이 16개의 코드워드만을 이용하여 벡터양자화를 수행한 경우에도 기존의 신호편의제거 알고리즘과 같은 바이어스를 추정하였으며, 최적으로 수렴하기까지 약 13회 정도의 추정이 필요하였다. 이는 기존 알고리즘에 비하여 6회 정도의 추정회수가 늘어났 것이다. 그러나 연산량을 계산하여 보면 기존 알고리즘의 경우 256개의 코드워드에 대하여 4번의 벡터양자화를 수행하였으며, 16개의 코드워드를 사용한 경우 16개의 코드워드에 대하여 13번의 바이어스추정을 수행하였으므로, 약 1/5정도로 연산량이 줄었음을 알 수 있다.

실제 인식을 측정실험에서는 새로운 바이어스를 추정하기 위해 벡터양자화를 수행할 때, 이전 추정에서 벡터양자화된 코드워드에 대하여는 벡터양자화 대상 코드워드에서 제외하였다. 이렇게 함으로써 바이어스 추정반복회수가 제한되어 있을 경우, 코드워드를 제한하는 알고리즘의 성능이 기존 알고리즘에 비하여 향상되었다.

그림 2는 각각 바이어스추정 반복회수와 제한된 코드워드수에 따른 인식률을 표와 그래프로 나타낸 것이다. 실험결과 32개의 코드워드를 사용했을 경우의 전반적인 성능이 가장 좋았으며, 8회의 바이어스 추정을 할 경우 연산량은 기존의 알고리즘에 비해 약 1/8감소하였다.

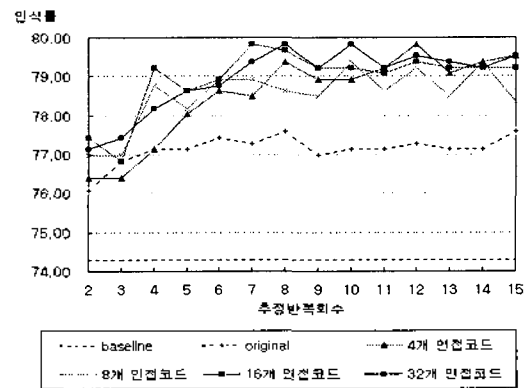


그림 2. 제한된 수의 코드워드를 이용한 신호편의제거 알고리즘의 인식률

Fig. 2. The recognition rate of the SBR algorithm using the limited number of codewords.

그림 2를 보면 추정 회수에 따라 인식률이 둔모양으로 변하고 있다. 이는 벡터양자화 수행시, 이전의 코드워드는 벡터양자화 대상에서 제외함에 따라 최적 바이어스값이 하나의 벡터로 수렴되지 못하고 진동하기 때문이다. 이러한 문제는 알고리즘을 약간 수정하여 해결할 수 있다.

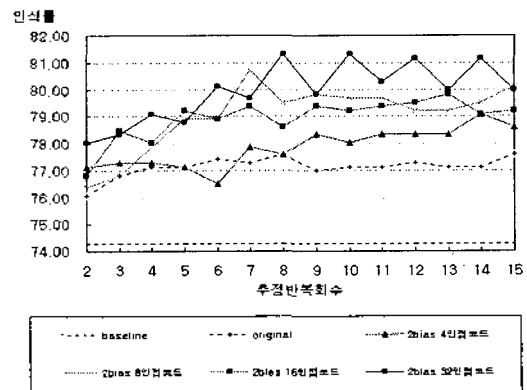


그림 3. 제안된 두 알고리즘을 동시에 적용한 신호편의제거 알고리즘의 인식률

Fig. 3. The recognition rate of the SBR algorithm incorporating the two proposed algorithms.

### 3.3.3 제안된 2가지 알고리즘을 동시에 적용한 신호편의 제거 알고리즘

앞에서 제안된 2가지 알고리즘을 결합하면 기존 알고리즘의 연산량을 줄이면서 성능을 최대한 향상시킬 수 있다. 두 가지 알고리즘을 결합할 경우 증가하는 오버헤드는 거의 없다. 메모리의 사용이 약간 증가할 뿐이며, 제한된 코드워드만을 사용하는 알고리즘에 비하여 연산량의 증가도 거의 무시할 수 있다.

그림 3에서 볼 수 있듯이 32개의 코드워드를 사용할 경우 가장 좋은 성능을 보였으며, 제안된 2개의 알고리즘을 결합하였을 경우 각각의 알고리즘을 단독으로 사용하였을 경우보다 훨씬 뛰어난 성능을 보였다.

## IV. 결 론

본 논문에서는 잡음에 오염된 음성을 보상하기 위하여 신호편의 제거 알고리즘을 기반으로 하여 새로운 보상 알고리즘을 제안하였으며, 전화음성 데이터베이스를 이용하여 제안된 알고리즘의 성능을 측정하였다.

입력음성의 SNR에 따라  $N$  개의 바이어스를 사용하는 알고리즘을 제안하여 기존 알고리즘의 성능을 좀더 향상시킬 수 있었다. 또한 기존의 전체 코드워드를 검색하는 벡터양자화 알고리즘을 개선하여 코드북의 일부 코드워드를 검색하도록 함으로써 연산량을 줄일 수 있었다.

제안된 두 가지 알고리즘을 결합하여 적용한 실험결과, 기존 알고리즘에 비하여 약 1/8정도 연산량을 감소시킬 수 있었으며, 보상 후의 인식률도 기존 신호편의 제거 알고리즘의 77.58%에서 81.32%로 향상되었다.

또한, 신호편의 제거 알고리즘을 음성 모델의 훈련에 이용할 경우 인식률을 좀더 향상시킬 수 있으며[2], 본 논문에서 제안한 알고리즘을 인식시스템의 훈련단계에서부터 적용하는 방법을 실험 중에 있다.

## 참 고 문 헌

1. S.J. Cox and J.S. Bridle, "Unsupervised speaker adaptation by probabilistic spectrum filtering," Proc. ICASSP, pp. 294-297, 1989.
2. M.G. Rahim and B.H. Juang, "Signal bias removal for robust telephone based speech recognition in adverse environments," Proc. ICASSP, pp. 445-448, 1994.
3. J.C. Junqua and J.P. Haton, Robustness in Automatic Speech Recognition, Kluwer Academic Publishers, Boston, 1996.
4. 제영철, 안정모, 임계중, 구명완, "잡음환경 및 채널왜곡에 강인한 ARS 용 전화음성 인식방식 연구," 한국음향학회 논문지, 제 16권, 제 2호, pp. 41-48, 1997.
5. A. Acero, Acoustical Environmental Robustness in Automatic Speech Recognition, Kluwer Academic Publishers, Boston, 1992.

▲임 계 중(Kye Jong Yim) 1970년 8월 21일생



1995년 2월: 홍익대학교 전자공학과 학사

1997년 2월: 홍익대학교 전자공학과 석사

1997년 1월~현재: 현대전자 근무 중.

※ 주관심분야: 음성인식, 디지털 신호처리, 컴퓨터 네트워크

▲계 영 철(Young Chul Kay) 1957년 12월 29일생



1980년 2월: 서울대학교 전자공학과 학사

1982년 2월: 한국과학기술원 전기 및 전자공학과 석사

1991년 5월: Univ. of Southern California, Electrical Eng. Ph.D.

1991년 9월~현재: 홍익대학교 전자전기공학부 부교수

※ 주관심분야: 디지털 신호처리, 음성 및 영상인식, 로봇 비전

▲구 명 완(Myoung Wan Koo)

현재: 한국통신 멀티미디어 연구소 음성언어 연구실 실장 (1999년 제 18권 3호 참조)

※ 주관심분야: 음성인식, 합성, 자동통역 전화 연구 및 실용화