

음성 인식 시스템의 화자 적응 성능 향상을 위한 코드북 설계

On Codebook Design to Improve Speaker Adaptation

양 태 영*, 신 원 호*, 김 원 구**, 윤 대 희*
Tae-Young Yang*, Won-Ho Shin*, Weon-Goo Kim**, Dae-Hee Youn*

요 약

본 논문에서는 반연속 HMM(semi-continuous Hidden Markov Model) 음성 인식 시스템에 적용되는 베이시안 화자 적응(Bayesian speaker adaptation)의 성능 향상을 위해 코드북 변환 알고리즘을 제안하였다.

기존 베이시안 화자 적응 알고리즘의 경우 새로운 화자의 특징 분포와 코드북 사전 밀도의 차이가 큰 경우 적응 데이터와 코드북간의 잘못된 대응 관계를 얻을 수 있으며, 기준(reference) 코드북에 필요 이상으로 많은 코드워드가 존재하는 경우 적용된 코드북에도 불필요한 코드워드들이 남아 인식 과정에 혼란을 줄 수 있다. 이 문제점을 해결하기 위하여 제안된 코드북 변환 알고리즘에서는 주파수 영역의 포먼트 정보를 이용하였다. 화자 적응을 수행하기 앞서 코드북의 캡스트럼으로부터 포먼트를 추출해 내고, 이들의 분포를 적응 화자의 포먼트 분포와 일치 되도록 변환시켜 주었다. 이 변환된 포먼트들로부터 다시 캡스트럼을 구하여 변환된 코드북을 얻고, 이를 화자 적응의 초기 코드북으로 사용하였다. 제안된 알고리즘을 이용하였을 경우 코드북과 적응 화자의 음성 간의 정확한 대응 관계를 찾을 수 있었고, 불필요한 코드워드들이 인식 과정에서 사용되지 않도록 변환되어 인식률이 향상되는 것을 실험을 통해 확인하였다.

Abstract

The purpose of this paper is to propose a method improving the performance of a semi-continuous hidden Markov model(SCHMM) speaker adaptation system which uses Bayesian parameter reestimation approach.

The performance of Bayesian speaker adaptation could be degraded in case that the features of a new speaker are severely different from those of a reference codebook. The excessive codewords of the reference codebook still remain after adaptation process, which cause confusion in recognition process. To solve such problems, the proposed method uses formant information which is extracted from the cepstral coefficients of the reference codebook and adaptation data. The reference codebook is adapted to represent the formant distribution of a new speaker and it is used for Bayesian speaker adaptation as an initial codebook. The proposed method provides accurate correspondence between reference codebook and adaptation data. It was observed that the excessive codewords were not selected during recognition process. The experimental results showed that the proposed method improved the recognition performance.

I. 서 론

음성 인식에 HMM(Hidden Markov Model)[1, 2]을 적용할 경우, HMM의 코드북중에서 가장 이상적인 코드북은 각 음소와 코드워드가 하나씩 일대일 대응되는 코드북일 것이다. 이 경우 단순히 코드북 양자화 과정만을 거처도 음성 신호에 담겨 있는 음소를 알 수 있는 경우지

*연세대학교 전자공학과
Dept. of Electronics Eng. Yonsei Univ.
**군산대학교 전기공학과
Dept. of Electrical Eng. Kunsan National Univ.
접수일자: 1995년 10월 13일

반 실제로 이것은 불가능하다. 그러므로 실제로 가능한 가장 적절한 코드워드의 수를 갖는 코드북은 각 음소와 음소들간의 변화를 나타내는데 필요한 최소한의 코드워드를 갖는 코드북이다. 코드북을 구성하는 코드워드들이 한 음소의 변화와 그 음소와 다른 음소간의 변화를 가장 잘 나타낼 수 있도록 적정수가 할당되었다면 최상의 인식을 얻을 수 있겠지만, 할당된 코드워드들의 수가 적당 수보다 적다면 구분력이 떨어지는 양자화 결과를 HMM에 주게 되어 인식률도 떨어질 것이고, 필요 이상으로 많다면 HMM의 인식 과정에 혼란을 가중시키는 결과가 되어 역시 인식률은 떨어질 것이다. 따라서 음성 인식에 적합한 코드북을 찾는 방법이 필요하다.

음성 인식의 화자 적응[3, 4, 5]에서도 역시 이와 같은 문제점이 생길 수 있다. 새로운 화자로부터 얻는 적응 데이터는 학습 데이터에 비해 훨씬 적은 분량이다. 이 불충분한 데이터로부터 새로운 화자의 음성 신호 분포를 정확히 알아내는 것은 불가능하며, HMM의 코드북과 새로운 화자의 음성 신호 분포에서 같은 음소들 간의 대응 관계를 정확히 찾아내어 적용된 코드북을 구하는 것도 어려운 일이다. 잘못된 대응관계에 의한 코드북 적용은 각 음소에 해당하는 코드워드들의 수를 적정수보다 크거나 작게 변화시켜 인식률을 떨어뜨리는 결과를 가져올 수 있으며, 만일 대응 관계를 적절히 찾아 적용된 코드북을 얻는다고 하더라도 원래의 기준 코드북이 최상의 코드북이 아니라면 최상의 인식률은 역시 얻을 수 없다.

본 논문에서는 이런 문제점을 극복하기 위하여 반연속 HMM(semi-continuous HMM)[1] 화자 적응 방법 중 하나인 베이시안(Baysian) 화자 적응[6]을 수행할 때, 그에 앞서 코드북 변환 과정을 추가함으로써 음성 인식에 적합한 코드북을 구하고 화자 적응 성능을 높일 수 있는 방법을 제안하였다. 제안된 코드북 변환 방법은 주파수 영역에서 각 음소의 특징을 잘 나타내고 있는 포먼트(formant) 정보를 이용하였다. 기준 코드북의 포먼트 분포를 새로운 화자 적응 데이터의 포먼트 분포에 맞도록 바꿔주어 화자 적응 과정에서 정확한 대응 관계를 찾고 각 음소에 해당하는 코드워드들이 적절한 수를 갖도록 하였다.

제 2장에서는 반연속 HMM의 베이시안 화자 적응에 대해 다루었고, 제 3장에서는 제안한 코드북 변환 방법에 대해서 설명하였다. 실험 내용과 결과는 제 4장에서 다루었고, 제 5장에서는 결론과 추후 연구과제를 제시하였다.

II. 반연속 HMM의 화자 적응

반연속 HMM에서는 관찰 확률 B 와 코드북의 평균 μ 과 분산 B 와 코드북의 평균 μ 과 분산 Σ_l^2 을 새로운 화자의 음성 특징에 맞도록 변환하여 화자 적응을 수행한다. 베이시안 화자 적응은 HMM의 학습 과정이 최대 유사도 추정(maximum likelihood estimation)[7]을 이용하는

것과는 달리 최대 결과치 추정(maximum a posteriori estimation: MAP)[8]을 이용한 알고리즘이다. MAP에서는 학습된 HMM 모델 뿐 사전 분포(priori distribution) $P_0(\lambda)$ 를 갖는 랜덤 변수(random variable)로 생각한다. 새로운 화자의 화자 적응을 위한 음성 데이터를 $Y = y_1, y_2, \dots, y_T$ 라고 할 때 이 화자에 알맞는 λ 를 얻기 위해서는

$$\frac{\partial}{\partial \lambda} P(\lambda|Y) = \frac{\partial}{\partial \lambda} \frac{P(Y|\lambda)P_0(\lambda)}{P(Y)} = 0 \quad (2-1)$$

이 되도록 λ 를 바꿔주어야 한다.

구체적인 화자 적응 과정은 다음과 같다. 코드북의 평균 μ 를 새로운 화자에 맞도록 변화시키기 위해서 μ 가 평균 ν 와 분산 τ^2 을 갖는 가우시안 분포를 이루고 있다고 생각한다. 새로운 화자의 화자 적응 데이터 Y 에 대해 코드북의 전체 L 개의 가우시안 분포 중 y_l 를 가장 잘 나타내는 l 번째의 가우시안 분포를 찾는 다음 l 번째 가우시안 분포에 해당된 모든 y_l 들에 대해 평균 m_l 와 분산 σ_{ly}^2 를 구한다. 새로운 코드북의 평균 $\hat{\mu}_l$ 은 다음과 같이 m_l 와 ν 의 가중평균으로 구할 수 있다.

$$\hat{\mu}_l = \frac{n_l \tau^2}{\tau_{ly}^2 + n_l \tau^2} m_{ly} + \frac{\sigma_{ly}^2}{\sigma_{ly}^2 + n_l \tau^2} \nu \quad 1 \leq l \leq L \quad (2-2)$$

여기서 n_l 은 가중치로 코드북의 l 번째 가우시안 분포가 선택된 y_l 의 갯수이다. 또한 새로운 코드북의 l 번째 분산 $\hat{\Sigma}_l^2$ 은

$$\hat{\Sigma}_l^2 = \begin{cases} \sigma_{ly}^2, & \text{if } \sigma_{ly}^2 > \sigma_{\min}^2 \\ \sigma_{\min}^2, & \text{if } \sigma_{ly}^2 < \sigma_{\min}^2 \end{cases} \quad (2-3)$$

로 구할 수 있으며 새로운 화자의 적응 데이터의 분산 σ_{ly}^2 로부터 얻어지는 적용된 코드북의 분산 $\hat{\Sigma}_l^2$ 은 부족한 적응 데이터 분량 때문에 지나치게 작은 값을 가질 수 있으므로, 이것을 막기 위해 분산의 최저값 σ_{\min}^2 을 준다.

관찰 확률 B 는 새로운 화자의 적응 데이터를 모델 λ 에 대해 비터비 분할(Viterbi segmentation)[9]을 통해 최적 상태 열을 구하고 각각의 상태에 대해 코드북의 l 번째 분포가 선택된 상대적인 횟수를 구해 주어진 모델의 관찰 확률 B 와의 가중 평균으로 적용된 관찰 확률 \hat{B} 를 구한다.

$$b_j^x(l) = \frac{\sum_{t=1}^T \delta(s_t = j) \delta(y_t \approx l)}{\sum_{i=1}^T \delta(s_t = j)} \quad (2-4)$$

$$\hat{b}_j(l) = \alpha b_j^x(l) + (1-\alpha)b_j(l) \quad 1 \leq j \leq N, 1 \leq l \leq L \quad (2-5)$$

여기서 $\delta(s_t = j) = \begin{cases} 1, & \text{if 시간 } t \text{에서 } \lambda \text{의 상태 } s_t = j \\ 0, & \text{그 외의 경우} \end{cases}$

$\delta(y_t \approx l) = \begin{cases} 1, & \text{if } y_t \text{에 대해 } l \text{번째 가우시안 분포가 선택} \\ 0, & \text{그 외의 경우} \end{cases}$

이때 α 는 학습가중치(learning weight)로써 적응 민감도를 결정하게 된다.

III. 코드북 변환

일반적으로 HMM의 코드북은 L 차원 특징 벡터의 평균 μ 와 분산 Σ^2 을 갖는 L 개의 코드워드로 이루어져 있으며, 본 논문에서는 14차 켈스트럼의 128의 코드워드를 갖는 코드북을 사용하였다. 이들 평균 μ 와 분산 Σ^2 은 반연속 HMM의 학습 과정을 통해 얻어지는데, 평균 μ 의 초기값은 LBG 알고리즘에 의해 구한 코드북을 사용하여 학습하였다.

음성 신호에서 모음은 자음보다 더 긴 시간동안 나타난다. 위와 같은 코드북에서는 시간적으로 긴 모음의 특징 벡터들은 자음의 특징 벡터들에 비해 더 밀집된 분포를 갖게 되고, 따라서 더 많은 코드워드가 할당될 것이다. 또한 화자 적응 과정의 경우, 적응 데이터가 충분한 분량이 아니기 때문에 새로운 화자의 음성 특징을 완벽히 나타낸다고는 볼 수 없고, 기존 코드북과의 정확한 대응 관계를 찾기는 어렵다. 반연속 HMM의 베이지안 적응방법의 경우 새로운 적응 데이터의 특징 벡터와 가장 가까운 코드북의 코드워드를 선택하여 그 코드워드를 새

로운 특징 벡터와 가중 평균으로 적응시켜가기 때문에 기존 코드북에 따라서 서로 다른 음소의 코드워드가 잘못 선택될 위험이 있으며, 특히 기존 코드북의 분포와 새로운 화자의 특징 벡터 분포 사이의 차이가 크면 클수록 이 위험성은 커진다.

따라서 본 논문에서는 새로운 화자에 대해 화자 적응 과정을 할 때 위와 같은 문제점을 해결할 수 있는 코드북 변환 방법을 제안하였다. 제안된 코드북 변환 알고리즘의 흐름은 그림 1과 같다.

3.1. 자음 모음 구별

본 논문에서는 클리핑 자기상관함수(clipping autocorrelation function)[10]를 이용하여 자음 구간과 모음 구간을 구별하였다.

먼저 샘플링된 음성 신호 $x(n)$ 을 저역 필터에 통과 시킨 후, 클리핑 문턱치 C_{thr} 를 찾아 이를 기준으로 신호를 클리핑하고 그 결과신호 $x_{clip}(n)$ 을 이용하여 자기상관함수를 계산하는 것이다.

$$x_{clip}(n) = \begin{cases} 1, & \text{if } x(n) > C_{thr} \\ -1, & \text{if } x(n) < -C_{thr} \\ 0, & \text{그 외의 경우} \end{cases}$$

C_{thr} : 클리핑문턱치(3-1)

이렇게 얻은 클리핑 신호 $x_{clip}(n)$ 의 정규화된 자기상관함수 $R_{clip}(\tau)$ 를 구한다. 이 $R_{clip}(\tau)$ 는 모음 구간에서는 피치(pitch)에 해당하는 곳에서 큰 값의 피크를 갖고 자음이나 묵음 구간에서는 피크가 발견되지 않는다. 따라서 이 피크의 존재 유무로 모음 구간을 구별할 수 있다[11].

3.2. 포먼트 추출

코드북과 적응 데이터는 14차의 켈스트럼이다. 이 켈스트럼에 제로 패딩(zero padding)을 거친 후 FFT를 수행하여 스펙트럼을 구한다.

이와 같은 과정을 거쳐 얻어지는 스펙트럼은 피치 성분이 모두 제거된 스무딩(smoothing)된 스펙트럼이므로 단순히 전후 값을 비교하여 피크가 되는 값을 찾는 방법으로도 포먼트를 비교적 정확히 찾을 수 있다.

3.3. 포먼트 분포 변환

사람의 음성 신호는 각 개인마다 서로 다른 특징을 갖고 있다. 음성 신호의 스펙트럼으로부터 얻는 포먼트들도 개인마다 다른 분포를 이루고 있다. 그러나 전체적인 분포는 다르지만 각 음소들 사이의 상대적인 포먼트 위치는 개인에 관계 없이 비슷한 위치에 놓이게 된다. 그러므로 서로 다른 사람의 포먼트들을 전체적인 분포만 일치하도록 변화시켜주는 것만으로도 각 음소들의 포먼트들은 같은 음소들끼리 비슷한 위치에 놓이게 되어 음소

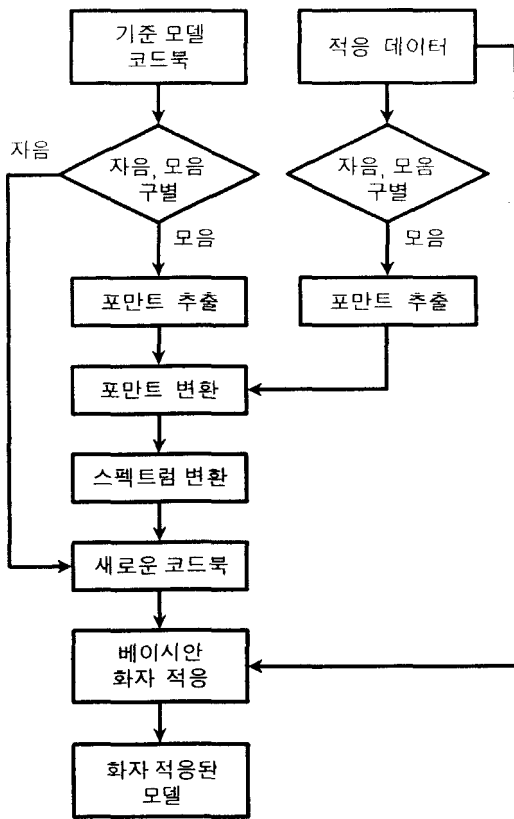


그림 1. 제안된 코드북 변환을 거친 베이지안 화자 적응 알고리즘의 플러드

Fig. 1. Block diagram of Bayesian speaker adaptation algorithm with proposed codebook adaptation

들간의 올바른 대응 관계를 갖도록 변화될 수 있다.

포먼트 분포 변환은 다음과 같은 과정을 반복적으로 수행하여 얻어진다.

step 1) 보음의 적응 데이터 중에서 임의로 한 프레임을 선택한다.

step 2) 선택된 적응 데이터 구간과 가장 유사한 포먼트를 갖고 있는 코드워드를 $K(iter)$ 개 찾는다.

step 3) V 개의 포먼트를 사용할때, 선택된 적응 데이터의 j 번째 포먼트를 f_j^a 라 하고 이 포먼트와 i 번째로 가까운 코드북 코드워드의 j 번째 포먼트를 f_{ij}^c 라고 하면, 적응된 코드북의 포먼트 f_{ij}^c 는 다음과 같이 구한다.

$$\hat{f}_{ij}^c = (1-\alpha)f_{ij}^a + \alpha f_j^c \quad 1 \leq i \leq K(iter), 1 \leq j \leq V \quad (3-2)$$

여기서 $K(iter)$: 적응되는 코드워드의 수

V : 적응되는 포먼트의 수

α : 학습 가중치, $0 < \alpha < 1$

학습 가중치(learning coefficient) α 는 적응 민감도를 결정하게 된다. 적응 데이터의 포먼트와 가장 가까운 것으로 선택된 코드워드의 수 $K(iter)$ 는 반복 회수에 대한 함수로써, 처음에는 코드북의 전체적인 분포가 적응 테

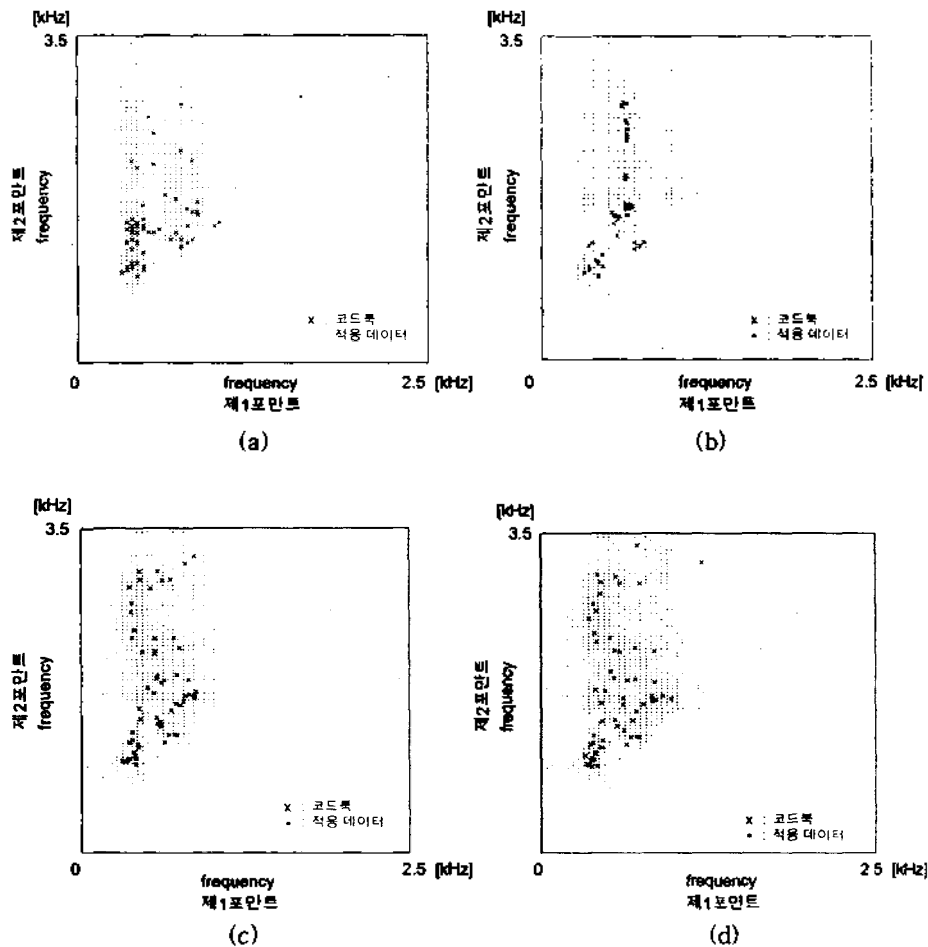


그림 2. 포먼트 분포 변화 과정

- (a) iteration = 0, $K = 30$
- (b) iteration = 200, $K = 10$
- (c) iteration = 1000, $K = 5$
- (d) iteration = 3000, $K = 1$

Fig. 2. Formant distribution adaptation procedure

- (a) iteration = 0, $K = 30$
- (b) iteration = 200, $K = 10$
- (c) iteration = 1000, $K = 5$
- (d) iteration = 3000, $K = 1$

이더의 분포와 비슷해질 수 있도록 $K(ite)$ 를 큰 값으로 설정하고, 포맷트 분포 변화 과정을 반복해 갈수록 점점 작은 값으로 바꿔주어 코드워드 하나하나가 적응 데이터의 각 음소들을 나타낼 수 있도록 해준다. 포맷트 분포 변화의 예들 그림 2에서 보았다.

3.4. 스펙트럼 변환

코드북의 코드워드들은 14차 켈스트럼들이다. 코드북의 포맷트들을 적응 화자의 포맷트 분포를 잘 나타낼 수 있도록 적응시켜준 후 이 적응된 포맷트 분포를 켈스트럼에 반영시켜 주어야 완전한 적응된 코드북을 얻을 수 있다.

포맷트 분포 변화 과정에서 얻은 적응된 포맷트를 사용하여 코드북 스펙트럼을 선형 변환(linear transform)시켜 준 후, 변환된 스펙트럼을 역FFT를 거쳐 새로운 화자에 적응된 켈스트럼을 얻게 된다.

IV. 실험 및 결과

4.1. 음성 신호 분석 및 데이터 베이스 구성

음성 신호 분석은 조용한 실험실 환경에서 발음한 음성 신호를 10kHz 샘플링하고 20ms를 분석 구간으로 하여 10ms마다 $1-0.95z^{-1}$ 의 프리엠퍼시스 필터를 거쳐 14차 LPC 계수를 구하고 이로부터 14차의 켈스트럼 계수를 얻어 이것을 특징 벡터로 사용하였다.

인식 대상 단어로는 우리말을 구성하고 있는 음소들을 발음 형태에 따라 분류하여 선택한 61개의 단어들[12]을 사용하였다.

4.2. 음성 인식 시스템의 구성

본 논문에서는 반연속 HMM의 화자 적응을 실험하기 위해 화자 종속 시스템, 화자 독립 시스템과 화자 적응 시스템을 구현하여 각각의 인식률을 비교하고 화자 적응 시 기존 알고리즘과 코드북 변환을 수행한 후 화자 적응을 하는 제안된 알고리즘의 인식률을 비교하였다.

위의 세 시스템을 학습시키기 위한 화자로는 화자 종속 시스템에서는 남성 화자 1명과 여성 화자 1명이 각각 10번씩 발음한 단어로 학습을 하였는데, 남성 화자와 여성 화자의 음성은 큰 차이가 있기 때문에 서로 다른 시스템으로 각각 구현하였다.

화자 독립 시스템은 남성 화자 7명과 여성 화자 5명이 각각 2번씩 발음한 단어로 학습을 하였다. 여기서는 남성 화자와 여성 화자를 구분하지 않고 같은 모델을 사용하였다.

인식률을 평가하기 위해서는 위의 두 시스템의 학습 과정에 참가하지 않은 남성 화자 2명과 여성 화자 2명이 3번씩 발음한 단어를 인식시켜 인식률을 평가하였다.

화자 적응 시스템은 화자 종속 시스템에 인식률 평가를 위한 테스트 화자가 1번씩 발음한 단어로 화자 적응을

수행한 후 역시 각각 3번씩 발음한 단어로 인식률을 평가하였다. 화자 적응은 기존의 방법과 제안한 방법의 두 가지를 수행하여 성능을 비교 평가하였다.

4.3. 결과 및 고찰

본 논문에서 구현한 인식 시스템의 성능을 평가하기 위하여 남성 화자 1명으로 학습된 인식 시스템과 여성 화자 1명으로 학습된 인식 시스템 두가지의 화자 종속 시스템의 학습 화자에 의한 인식률을 구하였다. 각각의 인식률은 다음 표 1과 같다.

표 1. 화자 종속 시스템의 인식률

Table 1. Recognition rates of speaker dependent system [%]

	남	여
화자 종속	97.3	96.7

4명의 테스트 화자에 의한 화자 종속 시스템과 화자 독립 시스템, 화자 적응 시스템의 인식률을 구하였다. 화자 독립 시스템의 인식률은 표 2와 같다. 화자 적응 시스템은 남성 화자와 여성 화자에 의해 학습된 두개의 화자 종속 시스템에 테스트 화자가 1번씩 발음한 단어를 적응 데이터로 하여 화자 적응을 수행한 후 인식률을 평가하였다. 여기서 기존의 베이시안 화자 적응에 의한 결과를 화자 적응 1이라 하였고, 제안된 코드북 변환을 수행한 후 베이시안 화자 적응한 결과를 화자 적응 2라 하였다. 그 결과를 표 3에서 보이고 있다.

표 2. 화자 독립 시스템의 인식률

Table 2. Recognition rates of speaker dependent system [%]

	남1	남2	여1	여2	평균
화자 독립	78.7	86.3	80.3	87.4	83.2

제안된 코드북 변환 과정을 수행한 후 베이시안 화자 적응을 한 화자 적응 2의 결과는 같은 분량의 적응 데이터를 이용하여 화자 적응을 했을 때 인식률이 평균 94.3%, 95.8%로 기존의 베이시안 화자 적응만을 수행한 화자 적응 1의 결과보다 3.6%정도 향상된 인식률을 보였다. 특히 코드북과 적응 화자의 음성 특징간의 좀더 정확한 대응 관계를 찾을 수 있는 방식이므로 인식 시스템의 학습에 참여한 화자의 음성 특징과 새로운 화자의 음성 특징 사이의 차이가 클수록 제안된 화자 적응 방법이 더 우수한 인식률을 보였다. 학습 화자와 테스트 화자의 성별이 다른 경우 화자 적응 2는 화자 적응 1에 비해 평균 6.6%의 인식률 향상이 있었다.

그림 3은 "마음"의 음성 신호 파형과 코드북 양자화 결과이다.

표 3. 화자 적응 시스템의 인식률
Table 3. Recognition rates of speaker adaptive system

[%]

	남성 화자에 의해 학습된 화자 독립 시스템					여성 화자에 의해 학습된 화자 독립 시스템				
	남1	남2	여1	여2	평균	남1	남2	남1	남2	평균
no adapt	60.1	74.3	43.2	54.1	57.9	56.3	45.9	63.9	68.6	58.7
화자 적응 1	95.1	92.9	87.4	91.8	91.8	91.3	89.1	89.6	93.9	91.0
화자 적응 2	96.7	94.0	92.3	94.0	94.3	97.3	96.2	93.4	96.2	95.8

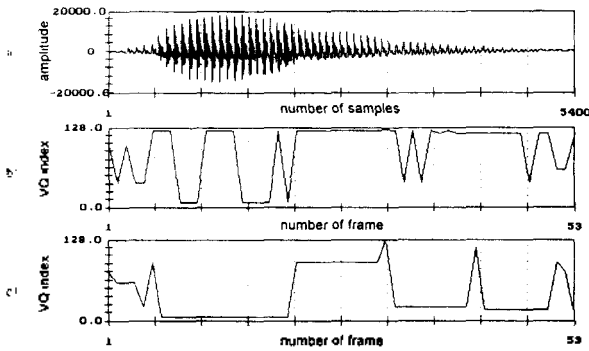


그림 3. 음성 신호 파형과 코드북 양자화 결과
(a) "마음"의 음성 신호
(b) 화자 적응 1에 의해 얻은 코드북 양자화 결과
(c) 화자 적응 2에 의해 얻은 코드북 양자화 결과

Fig. 3. Speech signal and vector quantization results
(a) speech signal of "Maeum"
(b) vector quantization result of speaker adaptation method 1
(c) vector quantization result of speaker adaptation method 2

128개의 코드워드로 구성되어 있는 코드북에 의한 양자화 결과인 그림 3을 보면 하나의 모음이 안정적으로 나타나는 구간에서 (b)의 경우에는 둘 이상의 양자화 결과가 나오는데 비해 (c)의 경우에는 하나의 코드워드가 대응된 것을 볼 수 있다. 이것은 제안된 코드북 변환 과정이 하나의 음소를 나타내는데 필요한 적절한 코드워드를 갖는 음성 인식에 적합한 코드북을 만들어주는 것을 의미한다. 따라서 제안된 변환 방법을 음성 인식 시스템의 학습 과정에 응용할 경우 더 높은 인식률을 얻을 수 있을 것이라 기대된다.

V. 결 론

본 논문에서는 반연속 HMM으로 단독음 인식 시스템을 구현하였고, 화자 종속 시스템과 화자 독립 시스템의 인식률을 화자 적응 시스템의 인식률과 비교하였으며, 화자 적응의 성능을 향상시키기 위한 방법으로 코드북 변환 알고리즘을 제안하였다.

제안된 코드북 변환 알고리즘은 주파수 영역의 포먼트를 이용한 코드북 적응 알고리즘으로써 하나의 적응 데이터를 이용하여 다수의 코드워드를 적응시킬 수 있으므로 화자 적응 데이터의 분량이 충분하지 못하기 때문에 발생하는 문제점을 극복할 수 있다. 이를 사용하여 화자 적응을 수행한 경우 기존의 화자 적응 알고리즘에 비해 인식률이 평균 3.6% 향상되었고, 특히 학습 화자와 테스트 화자의 성별이 다른 경우에는 평균 6.6%의 높은 인식률 향상을 보임을 실험을 통해 확인하였으며, HMM을 이용한 음성 인식 시스템의 학습 과정에 응용될 경우 인식률을 더 높일 수 있는 가능성을 제시하였다.

참 고 문 헌

1. L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE*, Vol. 77, No. 2, pp. 257-285, 1989.
2. L. R. Rabiner, B. H. Juang, "An Introduction to Hidden Markov Model," *IEEE ASSP Mag.*, Vol. 3, No. 1, pp. 4-16, 1986.
3. X. D. Huang, K. F. Lee, "On Speaker-Independent, Speaker Dependent, and Speaker- Adaptive Speech Recognition," *Proc. ICASSP*, pp. 877-880, May 1991.
4. K. Shikano, K. F. Lee, R. Reddy, "Speaker Adaptation through Vector Quantization," *Proc. ICASSP*, pp. 2643-2646, Apr. 1986.
5. S. Nakamura, K. Shikano, "A Comparative Study of Spectral Mapping for Speaker Adaptation," *Proc. ICASSP*, pp. 157-160, Apr. 1990.
6. C. -H. Lee, C. -H. Lin, B. -J. Huang, "A Study on Speaker Adaptation of Continuous Density HMM Parameters," *Proc. ICASSP*, pp. 145-148, Apr. 1990.
7. L. A. Liporace, "Maximum Likelihood Estimation for Multi-variate Observations of Markov Sources," *IEEE Trans. Informat. Theory*, Vol. 28, No. 5, pp. 729-734, 1982.
8. C. -H. Lee, J. -L. Gauvain, "Speaker Adaptation Based on MAP Estimation of HMM Parameters," *Proc. ICASSP*, pp. 558-561, Apr. 1993.
9. G. D. Forney, "The Viterbi Algorithm," *Proc. IEEE*, Vol. 61, pp. 268-278, Mar. 1973.
10. L. R. Rabiner, "On the Use of Autocorrelation Analy-

sis for Pitch Detection," *IEEE Trans. ASSP*, Vol. 25, No. 1, pp. 24-33, Feb. 1977.

- 11. Ji Hyun Nam, *Voice Personality Transformation*. Ph. D. Dissertation, Rensselaer Polytechnic Institute, Jan. 1991.
- 12. 한국 방송공사, "표준 한국어 발음 대사전," 어문각, 1993.

(약력)

▲양 태 영

1970년 3월 12일생



1993년 2월 : 연세대학교 전자공학과 졸업(공학사)

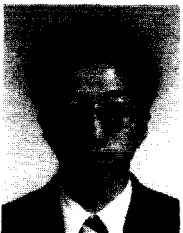
1995년 8월 : 연세대학교 대학원 전자공학과 졸업(공학석사)

1995년 9월~현재 : 연세대학교 대학원 전자공학과 박사과정

※주관심분야: 음성인식, 화자적응.

▲신 원 호

1967년 8월 24일생



1991년 2월 : 연세대학교 전자공학과 졸업(공학사)

1994년 2월 : 연세대학교 전자공학과 졸업(공학석사)

1994년 3월~현재 : 연세대학교 대학원 전자공학과 박사과정

※주관심분야: 음성인식, 잡음처리

▲김 원 구: 1994년 13권 1호 참조.

▲윤 대 회: 1994년 13권 1호 참조.