

에너지와 VQ를 이용한 음성 인식

Speech Recognition Using the Energy and VQ

황영수*

(Young-Soo Hwang)

요 약

본 논문은 음성인식과 화자 적응 방법들의 성능을 검토한 것이다. 본 논문에서 검토한 인식 방법은 에너지의 변화 상태와 VQ를 이용하여, 음성 인식을 수행하였으며, 화자에 따른 인식률 문제를 해결하기 위하여 화자 적응 방법(최대사후 확률추정법, 선형스펙트럼 추정법)을 검토하였다. 수행 결과 일반 VQ 방법보다 에너지 변화 상태를 이용한 인식 결과가 더 우수한 결과를 보였으며, 최대사후 확률 추정법을 적용할 경우, 2-3 %의 인식률 향상을 얻을 수 있었다.

ABSTRACT

In this paper, the performance of the speech recognition and speaker adaptation methods are studied.

The speech recognition using energy state and VQ(Vector Quantization) is suggested and the speaker adaptation methods (Maximum a posteriori probability estimation, linear spectrum estimation) are considered. The experimental results show that recognition ratio using energy state is 2-3 % better than that of general VQ.

Key words : FSVQ, energy state, speaker adaptation

1. 서 론

디지털 컴퓨터의 응용 기술과 반도체 기술 및 디지털 신호 처리 기술이 급격히 발전함에 따라 음성은 인간과 인간 사이의 의사 소통뿐만 아니라, 인간과 기계 사이의 의사 소통을 위한 매개체로서의 역할이 요구되고 있다. 인간의 가장 자연스러운 정보 교환 매체인 음성을 통하여 기계와 인간이 서로 정

확하게 정보를 전달하도록 하는 것을 목표로 한다.

음성인식 시스템은 1970년대 초부터 지금까지 활발히 연구되어 왔으며, 대표적인 인식 기법으로는 음성 발생시간 상에서의 패턴 정합에 의해 음성을 인식하는 DP (Dynamic Programming) 정합 방법, 인식 계산량과 메모리량을 적게 하기 위한 데이터 압축 기술을 이용한 벡터 양자화(Vector Quantization) 기법[1], Markov 모델의 확률적 추정에 의한

본 연구는 관동대학교 교내학술연구비 지원에 의해 수행되었음.

* 관동대학교 전자정보통신공학부 교수

† 논문접수일 : 2007년 11월 20일

기법을 도입한 HMM(Hidden Markov Model)과 음성의 인지 과정을 모델화한 인공 신경 회로망 등을 이용한 것들이 있으며, 현재는 위의 기법들을 서로 결합시켜 인식을 향상을 얻고자 노력하고 있다.

또한 대부분의 음성 인식 시스템은 화자 독립이거나 화자 종속 시스템으로 분류되며, 이 중 화자 독립 시스템은 사용자의 학습 단계를 요구하지 않으며, 많은 응용 분야에서 유용한 시스템이다. 그러나 사용 화자의 음향 특성의 변동 때문에 화자 종속 시스템보다 그 성능이 떨어지고 있는 실정이다. 그러므로 가장 이상적인 음성 인식 시스템은 사용함에 따라 사용자의 변화에 적응할 수 있는 시스템이다.

이와 같은 화자 특성 변동을 적응화하기 위하여, 음성 인식 시스템에 화자 적응 기능을 갖게 하는 방법에 대한 연구는 성대와 스펙트럼과 성도의 길이를 정규화하는 방법, 일부의 음소로부터 개인차에 적응하는 모든 음소의 스펙트럼을 추정하는 방법, 화자에 적응하는 표준 패턴의 집합을 선택하는 방법, 벡터 양자화에 의한 코드북의 매핑(mapping)방법 등이 있다

본 연구에서는 최대 사후 확률 추정법 (Maximum A Posteriori Probability Estimation)

[2], 선형 스펙트럼 추정 방법[3]들을 HMM에 적용시킨 결과[4]와 추후 비교 검토하여 더 나은 인식 방법을 모색하고자, 에너지와 VQ를 이용한 인식 방법에 적용하고자 한다. 선형 스펙트럼 추정 방법은 음향 특성을 추출한 후, 화자 특성을 제거시킨 방법이고, 최대 사후 확률추정법은 최대사후확률을 이용하여 최적 코드워드를 추출한 후, 화자 적응을 수행하였다.

II. 유한상태벡터양자화(FSVQ)와 화자 적응 방법

1. 유한상태벡터양자화(5-7)

1) 유한상태벡터양자화의 구조 및 원리

유한 상태 벡터 양자화는, 재생(reproduction) 코

드북과 복호기의 상태와의 차이에 따라, labeled state FSVQ(Finite State Vector Quantization) 구조와 labeled transition FSVQ구조로 구분할 수 있다.

본 논문에서 사용한 labeled state FSVQ의 부호기의 구조를 <그림 1>에 나타내었다.

<그림 1>에서 X_n 은 d차원의 입력 벡터이며, U_n 은 상태 코드북에 있는 코드워드를 나타내는 첨자(index)로서 복호기로 전송되는 채널 부호이다. 그리고 상태 코드북 C_s 란 X_n 을 부호화 하기 위해 슈퍼 코드북(super codebook) C 에서 선택된 N 개의 코드 워드로 구성된 코드북이다.

Labeled state FSVQ 부호화 방식은 상태 코드북(state codebook) C_s 만 주어지면, 일반적인 벡터 양자화 부호화 방식과 동일하다. 즉, 임의의 d차원 벡터가 입력되면, 상태 코드북 C_s 내의 코드워드 N 개 중 1개를 선택하여 전송하면, 일반적인 벡터 양자화 부호화 방식이 된다.

한편 labeled state FSVQ에 임의의 d차원 벡터가 입력되면, 상태 코드북 C_s 내의 코드워드 N 개 중 1개를 선택하여 현재의 상태로 하여 채널 부호를 전송하고, 다음 상태는 현재의 상태와 채널 부호 U (즉, C_s 내의 코드워드)에 의해 결정된다. 이것을 정의된 기호로 나타내면 다음과 같다.

$$\text{부호기 } \alpha : R^d * S \rightarrow N \quad (1)$$

$$\text{다음 상태 함수 } f : N * S \rightarrow S \quad (2)$$

식 (1)에서 R 은 벡터값, $N = \{0, 1, \dots, N-1\}$ 로, 구성된 채널 부호, $S = \{\sigma_i, i = 0, 1, \dots, K-1\}$ 로 구성된 상태 집합을 나타낸다. 여기에서 N 은 비트율을 결정하는 값으로서, 비트율은 $(\log_2 N/d)$ (bits/sample)이 된다.

한편 <그림 2>에 나타낸 labeled state FSVQ 복호기는 송신단과 동일한 상태 함수와 채널 부호를 이용하여 신호를 복호화 한다. 즉,

$$\text{복호기 } \beta : N * S \rightarrow R \quad (3)$$

이 된다. 단, 복호기의 초기상태는 부호기의 초기 상태와 같아야 한다. 이상의 내용을 정리하면, 다음과

같은 세 식에 의해 유한 상태 벡터 양자화를 규정할 수 있다.

$$U_n = \alpha(X_n, S_n) \quad (4)$$

$$S_{n+1} = f(U_n, S_n) \quad (5)$$

$$X_n = \beta(U_n, S_n) \quad (6)$$

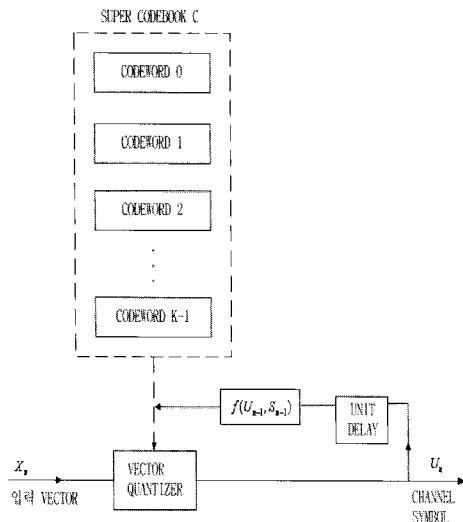
1) 유한상태벡터양자화의 인식 방법

미지 입력 음성의 분석 후, 이 음성의 프레임 수가 L개, i번째 프레임의 특징 벡터를 A_i , C_u 를 u 라는 단어의 코드북으로서 $C_u = \{c_{u1}, c_{u2}, \dots, c_{uk}\}$ 로 구성되어 있고, transition의 수(즉, 현재의 상태에서 변환 가능한 다음 상태의 수)를 N 이라고 할 때, 표준 패턴 내 단어 u 와 미지 음성의 비교는 다음과 같다.

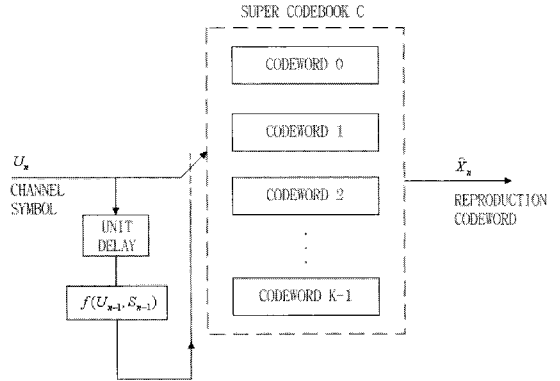
$$D_{u0} = \min d(A_0, c_{ui}), \quad i: \text{초기 상태} \quad (7)$$

여기에서 D_{u0} 는 입력 음성의 첫 번째 프레임과 초기 상태들과 비교하여 최소가 되는 왜곡 값이다.

만일, 초기 상태로 c_{uk} 가 선택 되었다고 하면,



<그림 1> labeled state FSVQ의 부호기
<Fig. 1> Encoder of labeled state FSVQ



<그림 2> labeled state FSVQ의 복호기
<Fig. 2> Decoder of labeled state FSVQ

$$D_{u1} = \min d(A_1, c_{uk}), \quad (8)$$

$$k = f(n_0, c_{uj}), f(n_1, c_{uj}), \dots, f(n_{N-1}, c_{uj})$$

이다. 여기에서 $f(\cdot)$ 는 다음 상태 함수이다.

일반적으로 입력의 i번째 프레임의 상태로 c_{uk} 가 선택되면, (i+1)번째의 최소 왜곡 $D_{u(i+1)}$ 은

$$D_{u(i+1)} = \min d(A_{i+1}, c_{uk}), \quad (9)$$

$$k = f(n_0, c_{uj}), f(n_1, c_{uj}), \dots, f(n_{N-1}, c_{uj})$$

이 된다.

그러므로 미지 입력 음성을 표준 패턴 단어 u 의 코드북 C_u 와 비교했을 때 평균 왜곡은 다음과 같다.

$$D_u = (\sum_{i=0}^{L-1} D_{ui}) / L \quad (10)$$

이러한 작업을 모든 표준 패턴들과 비교하여, D_u 중 최소가 되는 단어를 미지 입력 음성이라고 판정한다. 즉,

$$r = \min_u^{-1} D_u \quad (11)$$

위와 같은 방법으로 입력 음성을 인식할 경우, 계

산량은 transition의 수 N 에 비례하게 된다.

$$X_i^{(A)} = H^{(A)} L_i^A X_i^A \tag{10}$$

2. 화자 적응 방법

1) 최대사후확률추정을 이용한 화자적응

본 논문에서는 인식기로 벡터양자화[8]와 유한 상태 벡터 양자화를 사용하였기 때문에, 본 논문의 화자 적응 방법들을 벡터양자화와 유한 상태 벡터 양자화에 적용시켜 전개를 한다.

최대사후확률추정 방법[9,10]에서, 연속된 N 개의 샘플 벡터에 의한 평균 예측값은,

$$v'_N = \frac{av_0 + \sum_{i=1}^N X_i}{a+N} \tag{10}$$

으로 유도된다. 여기에서 X_i 는 인식기에 입력되는 샘플 벡터, a 는 상수, v_0 는 표준 모델의 평균 벡터이다.

본 연구에서는 식 (10)을 벡터양자화와 유한 상태 벡터 양자화에 적용시키기 위하여, v_0 를 표준 화자 음성을 이용하여 구성한 벡터양자화와 유한 상태 벡터 양자화의 격리 단어 모델 전체의 코드북내 코드워드로 설정하여 다음과 같은 식으로 변경시켰다.

$$v'_k = \frac{a v_{0k} + \sum_{i=1}^{N_k} X_i}{a + N_k} \tag{11}$$

식 (11)에서 X_i 는 미지 화자의 입력 벡터, v'_k 는 코드북내 k 번째 코드워드의 예측값, N_k 는 미지 화자의 연속된 입력 벡터중 k 번째 코드워드와 가장 유사도가 큰 입력 벡터의 갯수이다.

2) 음성 선형 특성을 이용한 화자 적응

임의의 화자 A 의 음성 특성을, 표준 패턴 화자 B 의 음성 스펙트럼의 선형 변화에 의해 다음과 같이 나타낼 수 있다.

여기에서 $H^{(A)}$ 는 A 화자의 음향학적 특성, $L_i^{(A)}$ 는 A 화자의 i 번째 음소 특성 변화식이다. 이와 같은 스펙트럼 변화의 양변을 log 화 한 후 선형 특성으로 변화시키면,

$$x_{i,t}^{(A)} = h^{(A)} l_{i,t}^{(A)} x_{i,t}^{(A)} \tag{12}$$

가 되고, 여기에서 $h^{(A)}$ 는 각 화자의 spectrum bias 라 할 수 있다. 그러므로,

$$h^{(A)} = \frac{1}{T} \sum_{t=1}^{T(A)} (x_t^{(A)} - U_{i^{(t)}}) \tag{13}$$

에서 구할 수 있다. 여기에서 $T(A)$ 는 화자 A가 발음한 음성의 프레임 수, $U_{i^{(t)}}$ 는 표준 패턴 코드워드중 화자 A가 시간 t 에서 발생한 $x_t^{(A)}$ 에 가장 유사도가 적합한 것이 된다.

이와 같이 구한 화자 A의 발생 특성 $h^{(A)}$ 를 A 화자의 음성에서 제거함으로써, 미지 화자와 표준 패턴 화자 상호간의 발생 특성을 제거할 수 있다

III. 에너지 상태를 이용한 음성인식방법

본 논문에서 사용한 격리 단어 인식 방법의 단계는 다음과 같다.

i) 에너지와 영교차율을 이용한 끝점 검출 알고리즘을 이용하여, 입력 격리 단어의 음성 구간을 추출해낸다.

ii) 음성 구간의 각 프레임에 대해 선형 예측 계수를 구하고, 현 프레임의 에너지, 두 프레임 이전의 에너지, 두 프레임 후의 에너지를 이용하여, 현 프레임이 ASCEND, FLAT, DESCEND 상태인지를 검토한다. 즉,

$$\frac{R_{i+2} - R_{i-2}}{R_i} > \Theta_1, \text{ -----} \rightarrow \text{ASCEND}$$

$$\Theta_2 \leq \frac{R_{i+2} - R_{i-2}}{R_i} \leq \Theta_1, \text{ -----} \rightarrow \text{FLAT}$$

$$\frac{R_{i+2} - R_{i-2}}{R_i} > \Theta_2, \text{ -----} \rightarrow \text{DESCEND}$$

R_i : i 번째 프레임의 0차 자기 상관 계수

Θ : 임계치 (본 논문에서는 1, -1 사용)

iii) 각 에너지 구간에 해당하는 코드북 내에서 표준 패턴 시 Histogram 방법으로 부터 구한, 전 프레임에 해당하는 다음 프레임의 비교할 코드워드에 대해, 왜곡을 측정 한다.

iv) 단계 iii)에서 구한 왜곡 값 중 가장 적은 왜곡값을 현 프레임의 왜곡값으로 하며, 이에 해당하는 코드워드를 다음 프레임 계산시 이용하기 위하여 저장한다. 즉,

$$D_{ui} = \min_k d(A_i, \ddot{E}, C_{uk}), \tag{14}$$

$$k = f(n, \ddot{E}, \ddot{E}, C_{uj}), \dots, f(n_{N-1}, \ddot{E}, \ddot{E}, C_{uj})$$

D_{ui} : u 라는 단어의 i 번째 프레임의 최소 왜곡

A_i : u 라는 단어의 i 번째 프레임의 특징 왜곡

C_{uj} : $(i+1)$ 번째 프레임의 최소 왜곡이 되는 코드 워드 (에너지 \ddot{E} 상태)

N : next 코드워드 수

v) 모든 표준 패턴들에 대해 구한 왜곡값 중 가장 적은 왜곡값을 갖는 표준 패턴을 입력 미지의 입력 단어라고 판정한다. 즉,

$$\text{미지의 입력 단어} = \min_u \frac{-1}{u} D_u$$

D_u : u 라는 단어의 평균 왜곡

위에서 서술한 에너지 파라미터를 이용한 유한 상

태 벡터 양자화 격리 단어 인식과정에서 사용한 코드북 구성 방법은 다음과 같다.

i) 표준 패턴을 구성시 사용되는 학습 데이터 각각에 대해 ASCEND, FLAT, DESCEND 부분으로 학습 데이터들의 프레임을 구분한다.

ii) 단계 i)에서 구분된 각 에너지 부분에 해당하는 프레임들에 대해 SPLIT 방법을 이용하여, 에너지 각 부분에 대한 코드북을 구성한다.

iii) 학습 데이터들에 대해 히스토그램(Histogram) 방법을 이용하여, 다음에 비교 할 코드워드들에 대한 확률분포를 구한다. 본 논문에서 사용한 히스토그램 방법은 아래 나타난 각 구간의 가능한 상태 변화 횟수를 누적시키는 방법을 사용하였다. 이때 각 구간의 가능한 상태 변화는 다음과 같다.

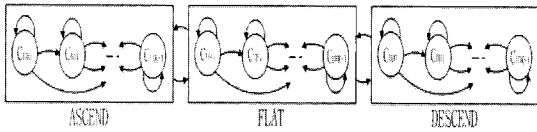
- ASCEND -----> FLAT,
- ASCEND -----> ASCEND
- FLAT -----> ASCEND,
- FLAT -----> FLAT
- FLAT -----> DESCEND,
- DESCEND -----> FLAT
- DESCEND -----> DESCEND

위의 가능한 변화 고려 상태시,

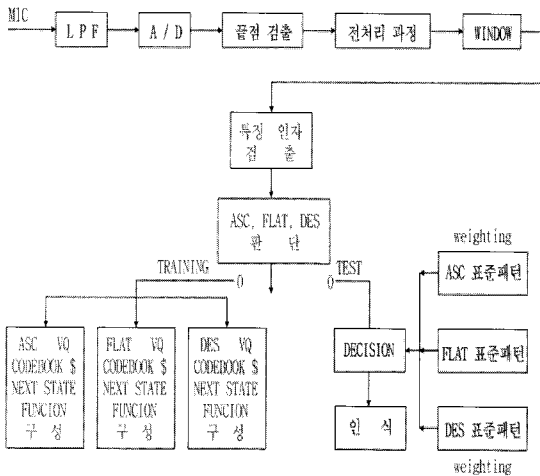
DESCEND -----> ASCEND,

ASCEND -----> DESCEND 에너지 변화 상태는 고려하지 않는다. 그 이유는 본 논문에서 사용된 데이터가, 잡음이 없는 상황에서 녹음된 음성을 이용하여기 때문에 음성이 상승 에너지 상태에서 갑자기 에너지가 감소하는 경우와 그 반대의 경우는 발생하지 않기 때문이다. 그러나 주변 상황이 잡음이 많이 발생하는 경우에는, 위의 두 개의 경우를 포함시켜야 할 것으로 생각된다.

<그림 3>에 에너지를 이용한 유한상태모델을 나타내었다.



<그림 3> 에너지를 이용한 유한 상태 모델
<Fig. 3> Finite state model using energy



<그림 4> 에너지 변화와 유한 상태 벡터 양자화를
이용한 음성 인식 시스템
<Fig. 4> Speech recognition system using
energy and FSVQ

IV. 실험 및 결과 고찰

1. 실험 데이터

실험에 사용된 데이터는 20대 남성 화자 9명으로 구성하였으며, 이 중 다섯명은 숫자(0-9)를 5번씩, 나머지 4명의 화자는 10번씩 반복 발음한 데이터로 구성하였으며, 각각 다른 시기에 발성한 데이터를 실험에 사용하였다. 기준 패턴을 구성하기 위하여 각 단어에 대해 네명의 화자가 열 번씩 반복 발음한 음성을 이용하였다.

데이터는 10KHz, 16bit로 샘플링(sampling)하였으며, 인식 파라미터로는 14차 선형예측계수와 자기

상관계수를 학습과 인식 등에 이용하였다.

2. 실험 결과

<표 1>은 코드워드 개수에 따른 벡터양자화의 인식 결과를 나타내었다.

<표 1>에 나타난 인식 결과를 살펴보면, 8개의 코드워드인 경우 66%, 24개의 코드워드인 경우 73.6%로 나타났으며, 코드워드 수와 관계없이 화자 3이 전체적으로 인식률이 저하된 것을 볼 수 있다.

이 결과는 표준패턴의 프레임 수가 상대적으로 많은 수를 갖기 때문에 코드워드 수가 많을수록 여러 화자의 특징을 갖는 것이라고 생각된다.

<표 1> 벡터 양자화를 이용한 인식 결과(화자별)
<Table 1> Recognition result using VQ

화 자	코드워드 8개	코드워드 24개
1	78%	76%
2	80%	78%
3	46%	64%
4	64%	76%
5	62%	74%
평 균	66%	71.6%

유한 상태 벡터 양자화의 next 코드워드 수는 3 개로 하였으며, conditional 히스토그램을 사용하였다.

<표 2>에 8개의 코드워드와 24개의 코드워드로 구성된 표준패턴의 인식 결과를 화자별로 나타내었다.

<표 2> 유한상태 벡터 양자화를 이용한 인식
결과(화자별)

<Table 2> Recognition result using FSVQ

화 자	코드워드 8개	코드워드 24개
1	80%	78%
2	80%	66%
3	60%	64%
4	74%	76%
5	62%	74%
평 균	71.2%	71.6%

인식 결과, 코드워드 8개와 24인 경우, 각각 71.2%, 71.6%의 인식률로 코드워드 수에 따른 인식률의 차이는 크게 나타나지 않은 결과를 보였다.

<표 3>은 에너지변화 상태와 FSVQ를 이용한 인식 결과를 나타내었다.

본 실험에 사용된 기준패턴은 각 숫자음에 대해 에너지 변화 상태식을 이용하여, 각 에너지 구간에 해당하는 프레임들로 표준패턴을 구성하였다. 즉, 각 숫자음에 대해 한 에너지 구간당 8개의 코드워드(각 숫자음에 대해 2개의 코드워드)로 구성된 코드북을 구성하였다. 유한상태 벡터양자화의 next 코드워드 수는 3개로 하였다. 또한 에너지 변화 상태식의 임계치는 각각 -1과 +1로 하여 음성의 프레임들을 구분하였다.

<표 4>에는 각각의 인식 방법에 화자 적응 방법을 적용시킨 결과를 나타내었다. 인식 결과 각각 인식 방법에 모두 화자 적응을 시킨 결과가 적응을 하지 않은 결과보다 더 나은 결과를 보여주었으며, 음성선형 특성을 이용한 화자 적응 방법보다 최대사후 확률 화자 적응 방법이 전반적으로 더 나은 인식 결과를 보여준다.

<표 3> 에너지 변화 상태를 이용한 인식 결과(화자별)

<Table 3> Recognition result using energy transition state

화 자	에너지 상태 이용		기존의 방법[1],[2]	
	VQ	FSVQ	VQ	FSVQ
1	76%	84%	78%	80%
2	94%	92%	80%	80%
3	90%	96%	46%	60%
4	98%	94%	64%	74%
5	80%	76%	62%	62%
평균	87.6%	88.4%	66%	71.2%

<표 4> 화자 적응 방법에 따른 인식 결과(화자별)
<Table 4> Recognition result using Speaker Adaptation

(a) 벡터양자화에 적용

화 자	MAPE	선형	MAPE+선형	적용안함
1	78%	77%	80%	76%
2	95%	94%	95%	94%
3	92%	90%	93%	90%
4	98%	98%	98%	98%
5	82%	81%	83%	80%
평균	89%	88%	89.8%	87.6%

(b) 유한상태벡터양자화에 적용벡터양자화에 적용

화 자	MAPE	선형	MAPE+선형	적용안함
1	86%	84%	87%	84%
2	92%	92%	92%	92%
3	96%	96%	96%	96%
4	94%	94%	94%	94%
5	79%	77%	80%	76%
평균	89.4%	88.6%	89.8%	88.4%

V. 결 론

본 논문에서는 음성 인식과 화자 적응 방법에 대해 연구를 수행하였다.

본 논문에서는 에너지 변화 상태를 벡터 양자화와 유한 상태 벡터 양자화에 적용한 결과, 인식 계산량은 벡터양자화보다 적었고, 유한 상태 벡터 양자화와 비슷한 결과를 얻었다. 그리고 인식률은 각각 벡터양자화 73.6%, 유한 상태 벡터 양자화 71.6%, 에너지 변화 상태를 이용한 벡터양자화 87.6%, 에너지 변화 상태를 이용한 유한 상태 벡터 양자화인 경우 88.4%로 나타났으며, 벡터 양자화, 유한 상태 벡터 양자화 모두 에너지 변화 상태를 이용한 결과가 사용하지 않은 결과보다 더 나은 인식률을 보였다.

또한 화자 적응 방법을 벡터 양자화 방법과 유한 상태 벡터 양자화 방법에 적용하여 실험을 수행하였

다. 실험 결과, 최대사후확률추정을 이용한 화자 적응 결과 벡터 양자화인 경우 89%, 선형 스펙트럼 추정 방법을 경우는 88%, 두 화자 적응 방법을 결합시킨 경우에는 89.8%로 화자 적응을 하지 않은 경우보다 2%정도의 인식을 향상을 얻을 수 있었다. 또한 같은 방법으로 유한 상태 벡터 양자화에 적용시켰을 경우, 최대사후확률추정을 이용한 화자 적응 결과 89.4%, 선형 스펙트럼 추정 방법을 경우는 88.6%, 두 화자 적응 방법을 결합시킨 경우에는 88.4%로 화자 적응을 하지 않은 경우보다 1.4%정도의 인식을 향상을 얻을 수 있었다. 이 결과는 만족할 만한 결과는 아니지만, 향후 데이터를 보완한 지속적인 연구와 주변 잡음에 따른 음성 인식을 검토, 그 대응 방법에 대해 연구를 진행할 예정이다.

참고문헌

- [1] Y. Linde, A. Buzo, and R. N. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Com.*, vol. 28, pp. 84-95, Jan. 1980.
- [2] M. Tonomura, T. Kosaka, and S. Matsunaga, "Speaker adaptation using maximum a posteriori probability estimation and data size dependent parameter smoothing," *전자정보통신공학회 논문지*, vol. J81-D-II, no. 3, pp. 465-471, Feb. 1998.
- [3] G. A. Carpenter, Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Neural Networks*, vol. 3, pp. 698-713, Apr. 1992.
- [4] Y. Hwang, "A study on the comparison of speaker adaptation methods," *Proc. ICSP'97*, pp. 591-594, Aug. 1997.
- [5] J. Foster, R. M. Gray, and M. O. Dunham, "Finite-state vector quantization for waveform coding," *IEEE Trans. Inform. Theory*, vol. 31, no. 3, pp. 348-359, May 1985.
- [6] M. O. Duham and R. M. Gray, "An algorithm for design of labeled transition finite-state vector quantizers," *IEEE Trans. Com.*, vol. 33, no. 1, pp. 83-89, Jan. 1985.
- [7] A. Haoui and D. Messerschmidt, "Predictive vector quantization," *Proc. Conf. Acoust., Speech, Signal Processing*, pp.10.10.1-10.10.4, Mar. 1984.
- [8] K. Shikano, K. F. Lee, and R. Reddy, "Speaker adaptation through vector quantization," *Proc. ICASSP 86*, vol. 1, pp. 2643-2646, Apr. 1986.
- [9] 황영수, "HMM을 이용한 화자 적응 방법에 관한 연구," *대한전자공학회강원지부 추계학술대회논문집*, 제10권, 제1호, pp. 34-38, Dec. 2005.
- [10] 정규철, 우수영, 진민호, 박용규, 유창동, "가중 훈련을 위한 화자 적응 시스템의 향상," *한국음향학회지*, 제22권, 제3호, pp. 188-193, 2003. 4.

저자소개



황 영 수 (Hwang, Young-Soo)

1982년 : 연세대학교 전자공학과 졸업(공학사)
 1984년 : 연세대학교대학원 전자공학과 졸업(공학석사)
 1990년 : 연세대학교대학원 전자공학과 졸업(공학박사)
 1989년-현재 : 관동대학교 전자정보통신공학부 교수