論文96-33B-1-12

은닉 마코프 모델을 이용한 음성 인식 시스템 설계

(Design of A Speech Recognition System using Hidden Markov Models)

李 喆 源 **, 林 寅 七 *

(Chul-Won Lee and In-Chil Lim)

요 약

본 논문에서는 이산 은닉 마코프 모델(Discrete Hidden Markov Model)을 이용한 연결 음성 인식에 관한 알고리듬 및 모델 토폴로지를 제안한다.

제안된 모델은 인식률과 인식할 수 있는 어휘를 고려하여 2 음소열 및 3 음소열 모델을 사용하며, 보다 정확한 음소 간의 세그멘테이션과 알고리듬의 수행 속도를 고려하여 2 음소열에서는 첫 번째 상태와 마지막 상태를 안정 상태, 나머지 상태는 천이 상태인 4 개의 상태를 갖도록 하고, 또한 3 음소열에서는 7 개의 상태를 갖도록 하며, 여기서 7 개의 상태는 3 개의 안정 상태와 4 개의 천이 상태를 갖도록 개선한다.

또한, 제안된 음성 인식 알고리듬은 인식 과정 내에서 음소의 발음 구간을 검출하도록 설계한다.

Abstract

This paper proposes an algorithm and a model topology for the connected speech recognition using Discrete Hidden Markov Models.

A proposed model uses diphone and triphone model which consider the recognition rate and recognisable vocabulary. Considering more exact inter- phoneme segmentation and execution speed of algorithm, 4 states have to exist in diphone model where the first state and the last state are keeping a steady state, the other states hold a transient state. 7 states have to exist in triphone model where 7 states are specified and improved to 3 steady states and 4 transition states.

Also, the proposed speech recognition algorithm is designed to detect the inter-phoneme segmentation during the recognition processing.

Ⅰ. 서 론

고도 정보화 사회로의 발전이 가속화되고 이에 따른 정보 수요의 급속한 증대로 인하여 정보 제공 시스템들이 다양화되면서 일반 사용자들에게 보다 자연스럽고 효율적인 음성 인터페이스의 필요성이 증대하고 있다. 이것은 음성을 이용한 사람과 기계 간의 통신이 가장 자연스럽고 편리한 것으로 생각되기 때문이다. 또한, 최근 고속 CPU를 탑재한 컴퓨터 기술 및 디지탈 신호 처리의 발전으로 대용량의 음성 신호를 실시간 고속 처리가 가능하여 음성 인터페이스를 위한 음성 인식의 중요성이 증대하고 있다.

* 正會員, 漢陽大學校 電子工學科
(Dept. of Elec. Eng., Hanyang Univ.)
** 正會員, 豆源工業專門大學 電子工學科
(dept. of Elec. Eng., Doowon Tech. Coll.)
接受日字:1995年5月10日, 수정완료일:1995年12月21日

이러한 음성 인식 기술은 크게 고립어 인식 및 연속 음성 인식, 그리고 연결 음성 인식 기법으로 나눌 수 있다. 고립어 인식은 미리 나누어진 몇 개의 음절로 구성된 단어나 문장 등의 음성 구간을 하나의 패턴으로 인식하는 것으로써 이는 화자의 발음에 제약이 따른다는 단점을 내포하고 있으며, 묵음과 단어 사이의 끝점 검출이 인식 과정 이전에 수행된다. 연속 음성 인식은 화자가 연속적으로 발음한 끝점이 알려지지 않은 음성 구간을 계속해서 인식을 수행하는 기법으로써 화자의 발음에 제약이 따르지 않지만 조음 결합 등에 의해서 인식률이 저하될 수 있다. 연결 음성 인식은 음소나 2 음소열 등 보다 작은 단위로 나누어 인식하여 단어나 문장을 재 구성하는 기법으로써 고립어 인식 및 연속 음성 인식의 단점을 보완한 기법이다. 또한, 연결 음성 인식 및 연속 음성 인식에 있어서는 음소 간의 구간 검출(segmentation)이 인식 과정과 병행해서 수행되어야 한다[3-4].

음성 인식 방법으로는 DTW (Dynamic Time Warping) 방법과 은닉 마코프 모델(Hidden Markov Model)에 의한 방법 등을 들 수 있다. DTW 방법은 음성 신호에서 유성음과 무성음의 발음 시간의 신축이 비선형적으로 나타난다는 문제를 해결하기 위하여 두 패턴 간의 비선형적인 신축에 의하여 시간 축을 정규화하는 방법으로써 비교적 높은 인식률을 나타내지만 많은 계산량과 한정된 어휘만을 인식할 수 있다는 단점이 있다. 은닉 마코프 모델은 음성 신호의 관측열과 모델의 상태열을 분리시킴으로써 보다 효율적으로 동적 계획법(dynamic programming)을 수행하도록 개선한 인식 방법으로 대 용량의 고속 음성 인식 시스템에 적합한 방식이다[5-7].

본 논문에서는 이산 은닉 마코프 모델을 이용한 연결 음성 인식에 관한 알고리듬 및 모델 토폴로지를 제안한다. 제안한 음성 인식 알고리듬은 인식 과정 내에서 음소의 발음 구간을 검출할 수 있도록 설계한다.

제안한 모델은 인식률과 인식할 수 있는 어휘를 고려하여 2 음소열 및 3 음소열 모델을 사용하며, 보다 정확한 음소 간의 세그멘테이션과 알고리듬의 수행 속도를 고려하여 2 음소열에서는 첫 번째 상태와 마지막 상태를 안정 상태, 나머지 상태는 천이 상태인 4 개의 상태를 갖도록 하며, 또한 3 음소열에서는 7 개의 상태를 갖도록 하고, 여기서 7 개의 상태는 3 개의 안정 상태와 4 개의 천이 상태를 갖도록 개선한다.

제안한 알고리듬과 모델 토폴로지의 효율성을 입증하기 위하여 단일 화자에 의해서 실제 발음된 문장으로 시뮬레이션을 수행하고 그에 대한 결과를 분석한다.

Ⅱ. 음성 인식 시스템

음성 인식은 이산적인 음운 정보가 연속적인 음성 신호로 변환되는 음성 합성의 역 과정, 즉 음성 신호로부터 음운 정보를 추출하는 과정을 말하며, 이러한 음성 인식 시스템의 일반적인 구조는 그림 1과 같다[2]

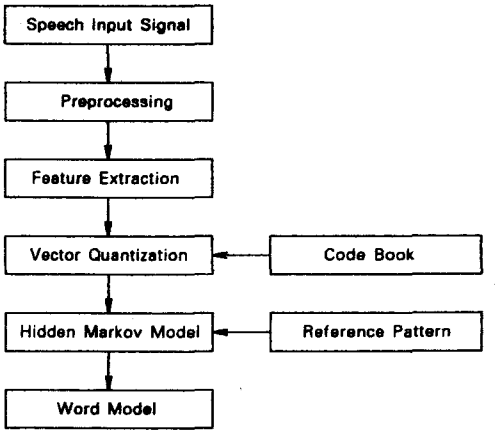


그림 1. 음성 인식 시스템 구조
Fig. 1. Structure of Speech Recognition System.

음성 인식 시스템은 데이타 획득 보드에 의해서 샘플링된 음성 신호을 입력으로 받아서 선형 예측 부호화(linear prediction coding;LPC)와 같은 음성 처리 기법을 통하여 관측열(observation sequence)을 얻어내는 전 처리 단계 및 관측열로 부터 통계적 모델을 통한 음성 패턴 매칭에 의해 음성을 인식하는 후 처리 단계로 크게 구분할 수 있다[2].

전 처리 단계에서는 음성 신호를 샘플링한 후 아날로그-디지탈 변환에 의하여 디지탈 신호를 얻는 데이타 획득 과정 및 성도관을 전극 필터(all-pole filter)로 가정하여 이에 대한 역 필터에 의해서 성도관 파라미터를 추출하는 특징 파라미터 추출 과정을 포함한다. 후 처리 단계에서는 전 처리 단계에서 추출된 특징 파라미터 시퀀스를 훈련된 통계적 모델이 대표하는 파라미터 시퀀스와의 오차를 distance 측정 기법을 사용하여 구한 후 가장 오차가 작은 모델이 나타내는 음성

정보를 인식된 값으로 결정하는 단계이다[2].

본 논문의 시뮬레이션에서는 데이타 획득 보드에서 획득한 음성 신호를 8 kHz의 주파수로 샘플당 16 비트의 분해능으로 샘플링하고, 음성 신호의 특징 파리미터 추출 시 계수값으로 0.97을 갖는 preemphasis 필터링을 수행 한 후, 20 msec (160 sample) 크기의 Hamming 창 함수와 Burg의 격자 필터(lattice filter)를 사용한 12 차 LPC 계수를 추출하여 이것으로 부터 12 차 LPC-Cepstrum 계수를 추출한다. 벡터 양자화 과정에서는 LPC-Cepstrum 계수를 training 벡터로 하여 코드 북의 크기를 64로 하며, modified k-means 알고리듬을 사용하여 클러스터링을 수행한다. 또한, 은닉 마코프 모델(Hidden Markov Model)을 이용한 패턴 매칭 단계에서는 이산 은닉 마코프 모델을 사용하여 6 개의 2 음소열 및 3 음소열를 훈련하며, 세그맨테이션에 필요한 최적 상태열을 구하기 위하여 Viterbi 알고리듬을 사용하고, Baum-Welch 알고리듬에 의해서 최적 상태열을 재추정한다.

Ⅲ. 음성 인식 시스템의 설계

1. 전 처리 및 특징 추출

음성 인식을 위한 전 처리 단계는 데이타 획득 보드를 통한 음성 신호를 8 kHz의 샘플링 주파수로 샘플당 16 비트의 분해능을 갖도록 하는 샘플링 과정과 이에 대한 12 차 LPC-Cepstrum 계수를 추출하는 과정으로 수행한다.

그림 2는 음성 신호를 획득하기 위한 데이타 획득 보드의 회로도를 나타낸 것이다.

Analog Device 사의 AD1848 SoundPort Stereo Codec과 IBM PC의 인터페이스는 DMA (Direct Memory Access) 방식을 적용하여 설계한다.

데이타 획득 보드에 의해서 얻어진 음성 신호를 8 kHz의 샘플링 주파수로 샘플링하고, 샘플된 신호는 주파수 영역에서의 다이나믹 레인지를 줄여 LPC 분석을 행할 때 수치적인 비 안정성을 줄이기 위하여 고주파성분의 에너지를 증가시키는 preemphasis 과정을 수행한다. preemphasis 필터는 식 (1)로 나타낼 수 있는 고대역 필터이며[2], 본 논문에서는 필터 계수 z_0를 0.97로 하여 수행한다.

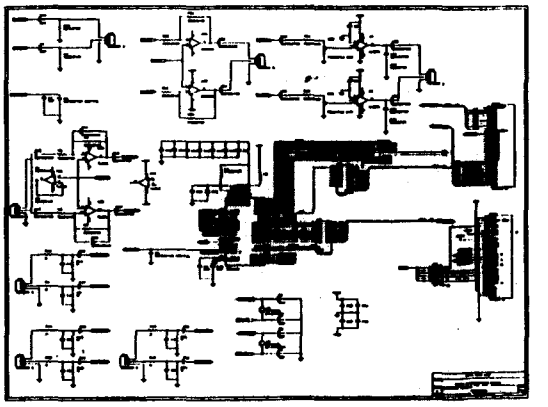


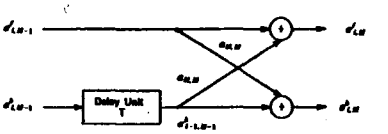
그림 2. 데이타 획득 보드
Fig. 2. Data Acquisition Board.

$$H(z)=1-z_0z^{-1},\quad z_0\simeq 1 \tag{1}$$

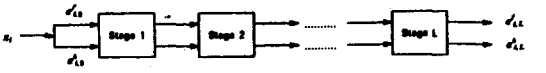
preemphasis 과정을 수행 한 후 20 msec 크기의 Hamming 창 함수와 격자 필터를 사용하여 12 차 LPC 계수를 추출한다. 사용한 Hamming 창 함수 $w(n)$을 식 (2)에 나타낸다[2].

$$w(n)=0.54\ -\ 0.46\cos\left(\frac{2\pi n}{N-1}\right) \tag{2}$$

12 차 LPC 계수를 추출하기 위한 격자 필터의 구조를 그림 3에 나타내며, 이를 수행하기 위한 Burg 알고리듬은 다음과 같다[1].



(a)



(b)

그림 3. 격자 필터의 구조
(a) 단위 격자 필터 (b) 다단계 격자 필터
Fig. 3. Structure of Lattice Filter.
(a) Single Stage (b) Cascade

▶ Burg 알고리듬 ◀

[단계 1] 각각의 필터 계수를 나타낼 인덱스를

$M = 0$으로 초기화하고, 각 변수는 아래와 같이 초기화한다.

$$a_{0,0} = 1, \quad e^f_{i,0} = e^b_{i,0} = x_i$$

여기서, $a_{0,0}$는 필터 계수의 초기값을 나타내고, $e^f_{i,0}$ 및 $e^b_{i,0}$는 각각 i번째 신호에서의 전 방향 예측 에러 및 후 방향 예측 에러의 초기값, x_i는 i번 째의 입력 시퀀스를 나타낸다.

[단계 2] M을 1 증가하고 아래의 식에 의해서 반사 계수를 추출한다.

$$a_{M,M} = \frac{-2\sum_{i=M+1}^{N} e^{b^*}_{i-1,M-1} e^f_{i-1,M-1}}{\sum_{i=M-1}^{N} (|e^b_{i-1,M-1}|^2 + |e^f_{i-1,M-1}|^2)}$$

여기서, $e^f_{i,j}$: j차에서 i 번째 신호의 전 방향 예측 에러

$e^b_{i,j}$: j차에서 i 번째 신호의 후 방향 예측 에러

$a_{M,M}$: M 차에서의 반사계수를 나타낸다.

[단계 3] 아래의 식에 의해서 필터 계수를 추출한다.

$$a_{M,M-m} = a_{M-1,M-m} + a_{M,M}\, a_{M-1,m}, \; m=1,\ldots,M-1$$

[단계 4] 아래의 식에 의해서 평균 오차를 계산한다.

$$\hat{P}_{M-1,N\ \min} = \frac{1}{2} D_{M-1}$$

$$D_{M-1} = -|e^b_{N,M-1}|^2 - |e^f_{N,M-1}|^2 + (1 - |a_{M-1,M-1}|^2)\, D_{M-2}$$

[단계 5] M이 주어진 필터의 차수보다 작으면 M을 1 증가하여 [단계 2]를 반복 수행한다.

LPC 계수를 구한 후 LPC 계수에서 식 (3) 및 식 (4)를 사용하여 12 차 LPC- Cepstrum 계수를 추출하며, 이를 음성 신호의 특징 파라미터라고 한다[2].

$$c_1 = a_1 \tag{3}$$

$$c_n = a_n + \sum_{m=1}^{n-1}(1 - \frac{m}{n}) a_m c_{n-m}, \quad 1 < n \le 12 \tag{4}$$

여기서, c_i 를 LPC-Cepstrum 계수라고 한다.

2. 벡터 양자화기

특징 파라미터 추출 과정에서 추출한 12차 LPC-Cepstrum 계수를 이용하여 벡터 양자화를 수행한다. 벡터 양자화기는 modified k-means 알고리듬[18]을 사용하여 설계하며, 그림 4에 modified k means 알고리듬의 흐름도를 나타낸다.

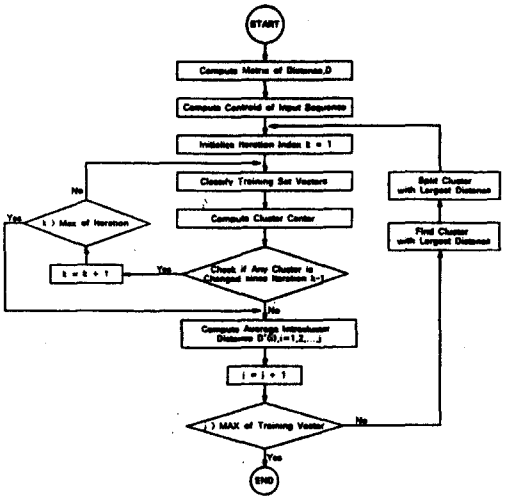


그림 4. 수정된 k-Means 알고리듬
Fig. 4. Modified k-Means Algorithm.

3. 은닉 마코프 모델(Hidden Markov Model)

은닉 마코프 모델(HMM)은 실제 관측열로 부터 확률적으로 다루어 지는 모델이며, 또한 변화하는 통계적 특성을 모델링하기 위해 마코프 프로세스를 이용한 것이다. HMM은 다음의 두 가지 가정을 고려하여 설계한다. 즉, 첫 번째 가정은 현재의 상태는 이전 상태에만 의존한다는 마코프 가정이고, 두 번째 가정은 출력 확률은 상태의 천이 과정과는 무관하게 특정 시간의 상태에만 의존한다는 것으로 이를 출력 독립 가정이라고 한다.

또한, 상태열은 감추어져(hidden) 있어서 단지 관측 가능한 관측열을 통해서만 추정할 수 있으므로 HMM은 이중적인 통계 모델이 된다. 모델의 각 감추어진 상태와 관측과의 관계는 관측 확률 분포에 의해 결정된다.

HMM을 이용하여 음성 인식을 수행하기 위해서 해결하여야 할 문제와 그에 대한 알고리듬은 아래와 같다[5-6].

첫 번째 문제는 관측열 $O = O_1 \ldots\ldots O_T$와 모델 $\lambda = (A, B, \pi)$가 주어질 때에 관측 확률 $P(O|\lambda)$를 구하는 것으로써 Forward 알고리듬 및 Backward 알고리듬에 의해서 재귀적으로 수행된다.

▶ Forward 알고리듬 ◀

[단계 1] Initialization

$$\alpha_1(i) = \pi_i b_i(o_1), \quad 1 \le i \le N$$

여기서, π_i : 초기 상태 분포

b_i : 관측열 확률 분포를 나타낸다.

[단계 2] Recursion

$\alpha_t(i)$를 모든 관측 시간과 모든 상태에 대해서 계산한다.

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij} \right] b_j(o_{t+1}),$$
$$1 \le t \le T-1, \quad 1 \le j \le N$$

[단계 3] Termination

최종 확률 값을 계산한다.

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

두 번째 문제는 주어진 모델과 관측열로 부터 최적의 상태열을 구하는 것으로써 Viterbi 알고리듬에 의해서 재귀적으로 수행된다.

▶ Viterbi 알고리듬 ◀

[단계 1] Initialization

$$\delta_1(i) = \pi_i b_i(o_1), \quad 1 \le i \le N$$
$$\psi_1(i) = 0$$

[단계 2] Recursion

모든 관측 시간과 상태에 대해서 다음을 계산한다.

$$\delta_t(j) = \max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}] b_j(o_t), \quad 2 \le t \le T, \ 1 \le j \le N$$
$$\psi_t(j) = \arg \max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}], \quad 1 \le j \le N$$

[단계 3] Termination

$$P^* = \max_{1 \le i \le N} [\delta_T(i)]$$
$$q_t^* = \arg \max_{1 \le i \le N} [\delta_T(i)]$$

[단계 4] Path (state sequence) back-tracking

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \ldots, 1$$

세 번째 문제는 주어진 관측열과 현재의 파라미터 값을 가지고 $P(O|\lambda)$가 최대가 되도록 새로운 파라미터 값을 결정하는 문제이다. 이는 Baum-Welch 재추정 알고리듬에 의해서 수행되며, 현재의 모델 $\lambda = (A, B, \pi)$에서 재 추정된 추정 모델 $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$은 식 (5)에서 식 (7)과 같이 재 추정할 수 있다.

$$\bar{\pi}_i = \frac{\alpha_0(i)\beta_0(i)}{\sum_{j=1}^{N} \alpha_T(j)} = \gamma_0(i) \tag{5}$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T} \alpha_{t-1}(i) a_{ij} b_j(o_t) \beta_t(j)}{\sum_{t=1}^{T} \alpha_{t-1}(i) \beta_{t-1}(i)} = \frac{\sum_{t=1}^{T} \xi_{t-1}(i,j)}{\sum_{t=1}^{T} \gamma_{t-1}(i)} \tag{6}$$

$$\bar{b}_i(k) = \frac{\sum_{t=1}^{T} \alpha_t(i) \beta_t(i) \delta(o_t, v_k)}{\sum_{t=1}^{T} \alpha_t(i) \beta_t(i)} = \frac{\sum_{t=1, \ s.t. \ o_t = v_k}^{T} \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(i)} \tag{7}$$

여기서, $\gamma_t(i)$는 관측열과 모델에 대해서 시간 t에서 상태가 i 일 확률을 나타내며, $\xi_t(i,j)$는 주어진 관측열과 모델에 대해서 시간 t에서의 상태가 i, 시간 t+1에서의 상태가 j일 확률을 나타낸다. 또한, $\alpha_t(i)$ 와 $\beta_t(i)$는 각각 시간 t에 대해서 상태 i의 forward 변수 및 backward 변수를 나타낸다.

4. 음성 인식 알고리듬

(1) 모델 토폴로지

음성 인식 시스템에서 사용되는 HMM의 모델 토폴로지는 Left-to-Right 모델을 사용하며, 이는 인식 시스템에 따라 약간의 변형을 하기도 한다.

본 논문에서 제안한 인식 시스템은 2 음소열 및 3 음소열 단위의 연결 음성이므로, 기존의 모델 토폴로지[7]를 사용하게 되면 2 음소열인 경우 6 개의 상태를 갖게 되고 각 상태로 부터의 천이는 13 혹은 15 가지가 된다. 또한, 3 음소열인 경우에는 9 개의 상태를 갖게 되고 각 상태로 부터의 천이는 22 혹은 24 가지가 된다. 따라서 제안한 모델은 2 음소열에 대해 4 개의 상태를 갖고 9 가지의 천이, 3 음소열에 대해서는 7 개의 상태를 갖고 18 가지의 천이를 할 수 있도록 설정한다.

또한, 연결 음성 인식에서 음소간 세그멘테이션 문제를 효율적으로 해결하기 위하여 모델의 각 상태에 대한 의미를 다음과 같이 정의한다.

즉, 2 음소열 및 3 음소열인 경우에서의 첫 번째 상태는 첫째 음소가 안정화된 정상 상태이고, 두 번째 상태는 첫째 음소에서 두 번째 음소로의 천이 상태로써

첫 번째 음소의 영향을 더 많이 받는 상태이며, 세 번째 상태는 두 번째 상태와 유사하며 단지 두 번째 음소의 영향을 더 강하게 받는 것이 다르다. 네 번째 상태는 두 번째 음소가 안정화된 상태를 나타낸 것이다.

또한, 3 음소열인 경우에서의 다섯 번째 상태는 두 번째 음소가 세번 째 음소로의 천이 상태를 나타낸 것이지만 두 번째 음소에 더 많은 영향을 받은 것이며, 여섯 번째 상태는 다섯 번째와 유사하며 단지 세 번째 음소의 영향을 더 많이 받은 것이다. 일곱 번째 상태는 세 번째 음소가 안정화된 상태를 나타낸 것이다.

그림 5는 제안한 모델의 토폴로지를 도시한 것이다.

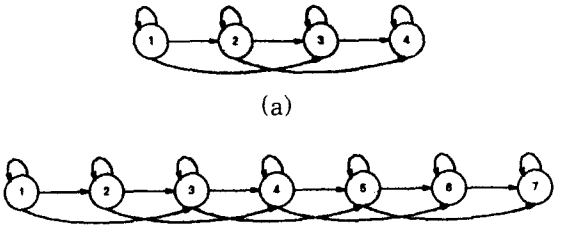


그림 5. 제안한 모델 토폴로지
(a) 2 음소열 모델 (b) 3 음소열 모델
Fig. 5. Proposed Model Topology.
(a) Diphone Model (b) Triphone Model

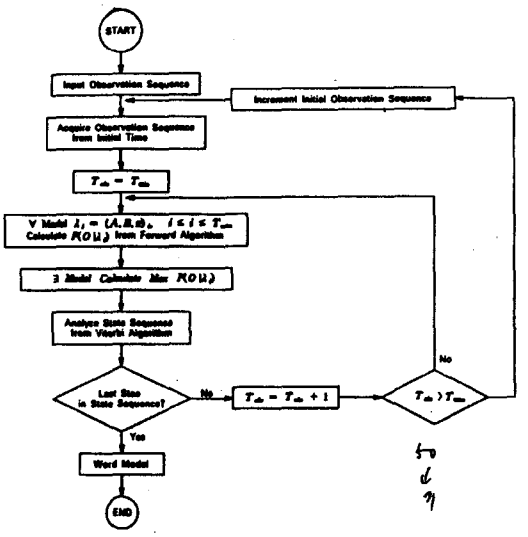


그림 6. 첫 음소 인식 알고리듬
Fig. 6. First Phoneme Recognition Algorithm.

(2) 인식 알고리듬의 구현

제안한 모델 토폴로지을 이용하여 인식 과정과 동시에 세그멘테이션을 수행할 수 있는 알고리듬을 구현한다.

최대 관측 시간 T_{max}는 음소간의 최대 천이 시간 구간으로 정의하며, 본 논문에서는 음소간의 최대 천이 시간을 300 msec로 설정하여, 프레임 윈도우가 20 msec 크기로 10 msec 씩 over-lapping 함으로써 $T_{max}=29$가 되도록 설정한다. 또한, T_{min} 은 최소 천이 시간 구간이며 $T_{min}=10$으로 설정한다.

임의의 크기를 갖는 관측열에 대해서 첫 음소를 인식하는 알고리듬을 그림 6에 나타낸다.

연결 음성 인식은 주어진 관측열로 부터 위와 같은 첫 음소 인식을 수행한 후에 얻은 상태열 중에서 마지막 상태에 해당하는 첫 번째 관측열로 부터 다시 반복적으로 수행한다.

Ⅳ. 실행 결과 및 분석

1. 음성 인식 수행 과정

본 논문에서 제안한 음성 인식 알고리듬의 수행 과정을 그림 7에 도시한다.

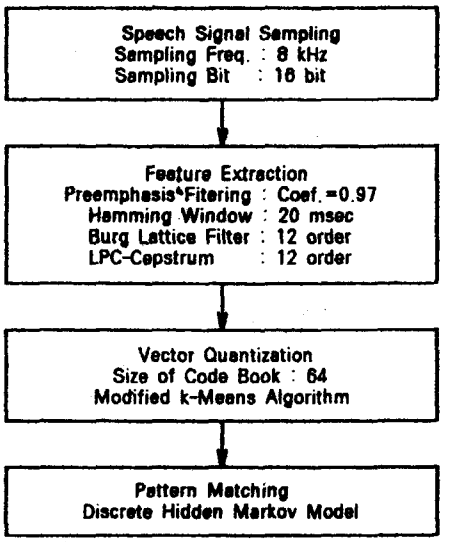


그림 7. 인식 알고리듬의 수행 과정
Fig. 7. Procedure of Recognition Algorithm.

2. 실행 결과

본 논문에서 사용된 음성 데이타는 단일 화자에 의해 반복된 것으로써 2개의 문장에 대해 각각 8 개의 2

음소열 및 3 음소열 모델을 설정하여 훈련시켰으며 3 개의 문장에 대해서 인식을 수행하며, 또한 인식 방법은 다음의 2 가지 방법에 의해서 수행한다. 즉, 첫 번째 방법은 첫 음소 인식 알고리듬을 수행하는 데 있어서 인식된 결과가 다음 음소의 인식에 영향을 주지 않도록 하며, 두 번째 방법은 첫 번째 음소의 인식 결과에 따른 유사도(likelihood) 값을 다음 음소 인식의 초기 확률로 사용하도록 한다.

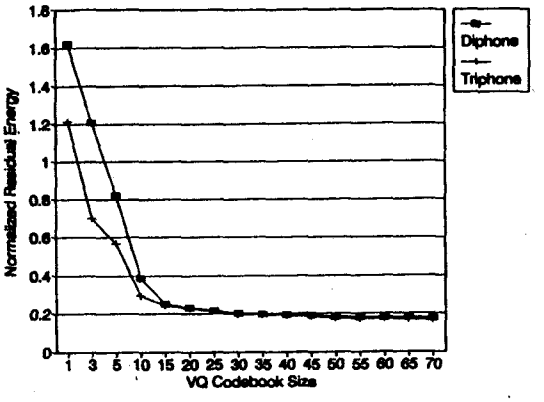


그림 8. 백터 양자화 코드북 크기에 따른 전체 왜곡의 상대적 크기
Fig. 8. Relative Size of Total Distortion for Vector Quantization Code Sizes.

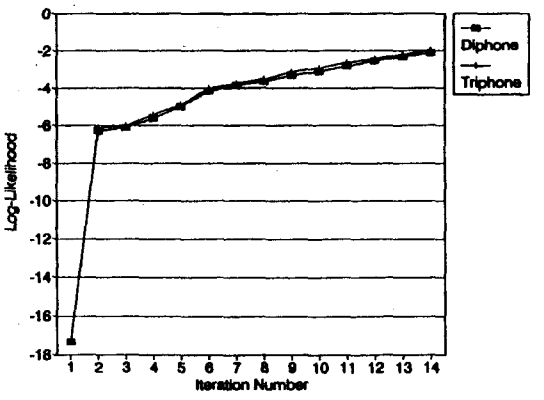


그림 9. Log-likelihood 증가 곡선
Fig. 9. Log-likelihood Incremental Curve.

Modified k-means 알고리듬에서의 코드 북 크기에 따른 양자화 성능을 그림 8에 도시하며, 또한 Baum-Welch 재 추정 알고리듬의 반복에 따른 log-likelihood의 증가 곡선을 그림 9에 나타낸다.

그림 8에서와 같이 코드 북 크기가 64 이상에서는 양자화 성능이 더 나아지지 않으므로 본 논문에서는 코드 북 크기를 64로 결정하여 벡터 양자화를 수행하였다.

제안한 인식 시스템에서 수행한 2 음소열 및 3 음소열에 대한 인식율을 표 1에 도시한다.

표 1. 수행 결과
Table 1. Execution Result.

수행 결과 구분		방법 I	방법 II
세그멘테이션 오류	2 음소열	1회 발생	없음
	3 음소열	없음	없음
오 인 식	2 음소열	1회 발생	없음
	3 음소열	없음	없음
인 식 율	2 음소열	92.3%	96.8%
	3 음소열	93.1%	97.2%

세그멘테이션 오류는 오인식에 의한 결과로써 산출될 수 있으며, 또한 이산 은닉 마코프 모델에 의한 유사도를 추정함으로써 세그멘테이션 과정을 수행할 수가 있다.

방법 I에서는 이전의 정보를 갖고 있지 않기 때문에 인식율이 다소 떨어지며, 방법 II는 첫번째 음소 인식 결과에 따른 유사도 값을 고려함으로써 모델 자체의 이전 정보를 갖고 있으므로 비교적 높은 인식률과 적은 세그멘테이션 오류를 나타낸다. 2 음소열에서 세그멘테이션 오류가 1회 발생되는 원인은 오인식에서 비롯된 것이다.

V. 결 론

본 논문에서는 이산 은닉 마코프 모델을 이용한 연결 음성 인식에 관한 알고리듬 및 모델 토폴로지를 제안하였다. 제안한 음성 인식 알고리듬은 이산 은닉 마코프 모델을 이용함으로써 인식 과정 내에서 음소의 발음 구간을 검출할 수 있도록 설계하였다.

제안한 모델은 인식률과 인식할 수 있는 어휘를 고려하여 2 음소열 및 3 음소열 모델을 사용하였으며, 보다 정확한 음소 간의 세그멘테이션과 알고리듬의 수행 속도를 고려하여 2 음소열에서는 첫 번째 상태와 마지막 상태를 안정 상태, 나머지 상태는 천이 상태인 4 개의 상태를 갖도록 하였고, 또한 3 음소열에서는 7 개의 상태를 갖도록 하였으며, 여기서 7 개의 상태는

3 개의 안정 상태와 4 개의 천이 상태를 갖도록 개선하였다.

결과적으로 적은 상태를 갖는 단순화된 모델 토폴로지와 알고리듬에 의해 인식 수행 속도의 향상을 기대할 수 있으며, 또한 인식 과정과 병행하여 세그멘테이션을 수행함으로써 세그멘테이션 오류도 줄일 수 있었고, 오인식에서 영향을 받는 세그멘테이션 오류가 다음 음소의 인식에 큰 영향을 주지 않음을 확인할 수 있었다.

본 논문에서 제안된 알고리듬은 음성 인식에 있어서 음소 간의 경계를 미리 나눌 수 없는 연속 음성 인식이나 연결 음성 인식에서 세그멘테이션 문제 해결에 도움을 줄 수 있을 것으로 기대된다.

앞으로의 연구 과제는 본 논문에서 제안된 알고리듬을 연속 은닉 마코프 모델(Continuous HMM) 및 반연속 은닉 마코프 모델(Semi-Continuous HMM)에 적용하여 보다 높은 인식률을 얻는 데에 있으며, 또한 불특정 화자에 대한 음성 인식 시스템의 설계에 있다.

참 고 문 헌

[1] A.A. Giordano and F.M. Hsu, *Least Square Estimation with Application Digital Signal Processing*, John Wiley & Sons, 1985.

[2] L.Rabiner and B.H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993.

[3] J.S. Bridle, "Stochastic Models and Template Matching : Some Important Relationships between Two Apparently Different Techniques for Automatic Speech Recognition", Institute of Acoustic Autumn Conf., 1984.

[4] J.S. Bridle, M.D. Brown and R.M. Chamberlain, "An Algorithm for Connected Word Recognition," *Proc. ICASSP-82*, pp.899-902, 1982.

[5] L.Rabiner and B.H. Juang, "An Introduction to Hidden Markov Models," *IEEE ASSP Magazine*, pp.4-16, Jan. 1986.

[6] A.B. Poritz, "Hidden Markov Models : A Guided Tour," *Proc. ICASSP-88*, pp. 7-13, 1988.

[7] B.J. Juang, "On the Hidden Markov Model and Dynamic Time Warping for Speech Recognition - A Unified View," *AT&T Bell Lab. Tech. Jour.* Vol.63, pp.1213-1243, 1984.

[8] J.G. Wilpon and L.R.Rabiner, "A Modified K-Means Clustering Algorithm for Use in Isolated Word Recognition," *IEEE Transaction on ASSP*, Vol. ASSP-33, No.3, Jun. 1985.

[9] G.D.Forney, "The Viterbi Algorithm," *Proc. of the IEEE*, Vol.61, No.3, Mar. 1973.

[10] -, "Parallel-Port 16 Bit SoundPort Stereo Codec : AD1848," Analog Devices Data Book.

저 자 소 개

李 喆 源(正會員) 第 28卷 B編 第 6號 參照
현재 한양대학교 대학원 박사과정 재학중. 현재 두원공업전문대학 전자과 전임 강사

林 寅 七(正會員) 第 28卷 第 4號 參照
현재 한양대학교 전자공학과 교수