

분할 확률 모델을 이용한 한국어 고립단어 인식

(Isolated Word Recognition Using Segment Probability Model)

金 秦 永*, 成 宏 模**

(Jin Young Kim and Koeng Mo Sung)

要 約

본 논문에서는 고립단어를 인식하기 위한 새로운 방법으로서 분할 확률 모델을 제시하고 인식 실험을 통하여 그 성능을 보였다. 분할 확률 모델은 통계적 모델로서 분할과 각 분절의 모델링의 두 과정을 이루어져 있다. 즉 입력된 단어를 임의의 개수로 분할한 뒤 각 분절을 벡터 양자화기를 이용 분절의 확률 분포에 따른 관찰 확률을 구하는 방법이다. 위의 방법을 인식실험을 통하여 기존의 방법인 패턴정합의 방법, Hidden Markov 모델과 비교하였다. 실험의 결과로써 제안한 방법이 인식률이나 계산량의 측면에서 기존의 모델보다 우수함을 확인할 수 있었다.

Abstract

In this paper, a new model for isolated word recognition called segment probability model is proposed. The proposed model is composed of two procedures of segmentation and modelling each segment. Therefore the spoken word is divided into arbitrary segments and observation probability in each segments is obtained using vector quantization.

The proposed model is compared with pattern matching method and hidden Markov model by recognition experiment. The experimental results show that the proposed model is better than existing methods in terms of recognition rate and calculation amounts.

I. 서 론

컴퓨터의 발달과 함께 인간은 기계와의 대화(communication)를 위해 언어가 가장 자연스러운 방법으로 생각하게 되어, 사람의 음성을 인지하는 시스템을 고안하고자 노력하게 되었다. 따라서 기계에 의한 음성의 인지, 즉 음성인식 시스템은 여러 사람의 관심의 대상이 되어 1960년대 후반부터 활발한 연구가 진행되어 왔다. 그 결과 음성인식을 위한 여러 방법들이 제시되었고 실용화되어 사용되어지고 있다.

특히 IBM에서는 2만단어 실시간 인식기를 만들어 실험하고 있는 실정이다.¹⁾ 음성인식은 연속음인식, 고립단어인식 그리고 화자인식의 세 분야로 나눌수 있는데, 그 중 한 분야인 화자독립 고립단어 인식은 구별적으로 발음되어진 단어를 인식하는 것으로서 음성인식의 가장 기본적인 분야이다. 지금까지 고립단어 인식을 위한 여러가지의 방법이 제시되어 왔으나 크게 패턴정합과 통계적 모델의 두 방법으로 분류할 수 있다. 패턴 정합의 방법은 인식하고자 하는 각각의 단어에 대하여 기준패턴을 만들고, 입력된 단어를 이것들과 거리를 계산하여 가장 비슷한 단어로 인식하는 시스템으로 그 인식률은 높다. 그러나 dynamic time warping(DTW)과 같은 복잡한 알고리

*準會員, **正會員, 서울대학교 電子工學科
(Dept. of Elec. Eng., Seoul Nat'l Univ.)
接受日字: 1988年 10月 6日

음을 이용함으로써 계산량이 커 많은 단어를 실시간 인식하기에는 어려운 점이 있다.¹²⁻⁹⁾ 따라서 1980년 이후 계산량을 줄이기 위한 방법들이 연구되어 왔는데 주로 통계적 모델을 생각하게 되었다. 통계적 모델에 의한 인식은 hidden markov model(HMM)과 같이 각 단어에 대해 어떤 모델을 새우고 그 모델에 기준하여 확률을 구하는 방법으로 계산량을 많이 줄일 수 있으나 인식률이 약간 떨어지는 단점이 있다.¹⁰⁻¹²⁾ 또한 HMM은 학습시에 Baum-Welsh 알고리즘 또는 forward-backward reestimation과 같은 복잡한 과정을 이용하고 인식시에도 Viterbi 알고리즘¹³⁾을 이용하고 있어 복잡한 계산을 필요로 하는 문제가 있다. 따라서 학습과 인식시 간단한 과정을 필요로 하는 고립단어 인식모델이 필요하다고 하겠다.

본 논문에서는 이러한 목적을 위하여 분할확률 모델이란 새로운 인식방법을 제시하고, 이 모델을 이용한 고립단어인식을 인식률과 계산량의 면에서 기존의 방법과 비교 검토하고자 한다.

II. 분할 확률 모델

1. 분할확률 모델의 배경

인식시에 계산량이 많이 감소되는 HMM은 인식대상 고립단어를 임의의 개수의 상태(state)로 나누어 그 각각의 상태에서의 관측 값의 확률과 상태 간의 천이확률을 갖는 Markov source로 표시하는 알고리즘이다. 여기서 각 상태는 특정 스펙트럼이 그 상태에서 관측될 확률밀도 함수(probability density function)를 갖게 되는데 인식시에는 관측된 스펙트럼 열이 각 상태를 어떠한 확률과정을 통해 나타내 있는가를 계산하게 된다. 한편 각 상태는 학습시에 Baum-Welsh와 같은 최적화 알고리즘을 이용한 최적화된 상태이므로 관측된 스펙트럼이 어느 상태에 속한다는 것은 그 스펙트럼이 어느 상태에서 관측 확률이 가장 큰 스펙트럼과 거리가 가장 가깝다는 것을 나타낸다. 그런데 고립단어라는 것은 시간에 따른 서로 다른 음소의 열이라 할 수 있으므로 각 상태가 특정 음소로 나타낸다고 생각할 때 각 상태간의 천이확률이 없어도 음소의 변화에 따른 스펙트럼의 변화는 자연적으로 상태를 옮겨놓게 된다. 즉 고립단어의 인식에 있어서 상태간의 천이확률은 큰 의미를 가지지 못하는 것으로 생각되며 각 상태의 길고 짧은만을 보여주는 것이다. 따라서 중요한 것은 상태천이 확률이 아니라 각 상태가 나타내는 스펙트럼 관측확률 분포이며 이에 따라 각 상태를 어떠한 확률로 나타냈는가 알 것이다. 위의 가정을 확인하기 위

하여 left-to-right 모델을 이용한 경우에 대하여 한국어 숫자음을 대상으로 하여 인식실험을 하였다. 인식실험은 50명의 화자에 대해 상태 개수 4개, codeword 개수 32개에 대하여 수행하였다. 그 실험한 결과 상태 천이확률을 고려할 때는 93.6%, 고려하지 않을 때 91%의 인식률을 보여 다소 인식률의 감소는 보였지만 큰 차이가 없음을 알 수 있었다. 따라서 위의 실험 결과를 생각하여 볼 때, 상태 천이확률을 무시한채로 상태를 잘 구할 수 있다면 HMM과 같은 복잡한 학습과 인식과정 Baum-Welsh, Viterbi 알고리즘을 이용하지 않는 간단한 모델이 가능함을 알 수 있다. 이를 위해서 입력된 단어를 일정한 법칙에 의하여 시간에 따라 결정적(deterministic)으로 나누는 방법을 찾으면 되겠고 이렇게 나누어진 상태의 관측확률을 구하면 되겠다. 이러한 모델링을 분할 확률 모델이라 부르기로 하고, 여기서 나누어진 각 state를 분절(segment)이라 부르기로 한다.

2. 분할과정

주어진 단어에서 임의의 개수의 분절(준정상상태; quasi-stationary)을 잘 찾는다는 것은 그리 쉬운 일이 아니다. 그런데 각 단어에서 임의의 개수의 분절을 찾는다는 것은 바꾸어 말하면 단어를 임의의 개수로 분할(segmentation)한다는 것을 의미한다. 분할은 연속음을 인식하고자할 때 필수적인 전처리 과정이므로 여러사람의 관심 대상이 되어 많은 방법들이 제시되어 왔으나, 대체로 분절을 위한 임계값이 필요하고 이 값에 따라 분절의 개수가 달라지게 된다. 따라서 임의의 고정된 개수의 분절을 구하고자 할 때는 적합하지 않다. 고립단어를 임의의 개수로 나누는 분할 방법에는 여러가지의 기준이 있을 수 있고 그에 따라 여러가지 방법이 있을 수 있다. 가장 간단한 분할방법으로는 입력단어를 일정한 시간간격으로 분할하는 방법을 생각할 수 있겠다. 그러나 이러한 방법은 스펙트럼의 변화를 고려하지 않은 것이기 때문에 각 분절에 여러 정상상태가 포함될 우려가 많아 바람직하지 못한 것이다. 그래서 스펙트럼의 변화를 고려한 방법을 생각해야 되는데 그 중에 궤적분할(trace segmentation)을 이용할 수 있겠다. 왜냐하면 궤적분할은 음성의 스펙트럼 열이 정상상태에서는 그 스펙트럼 변화량이 작다는 사실을 이용했기 때문이다. 만약 음성의 스펙트럼열이 $A_1, A_2, A_3, \dots, A_k$ 이라면 전체 스펙트럼 변화량을 C라고 할 때 C는 아래와 같다.

$$C = \sum_{i=1}^k d(A_i - A_{i+1}) \quad (1)$$

단, $d(,)$ 는 거리함수

위의 식에서 C는 각 이웃 스펙트럼의 거리의 합이 되는데, 만약 서로 다른 정상상태의 경계에 해당하는 부분이라면 스펙트럼의 변화량이 커 C값을 좌우하는 요소가 된다. 따라서 M개의 분절을 구하고자 한다면 국부적 스펙트럼열이 C/M이 되도록 잘라 M개의 분절을 구할 수 있다. 그런데 제적분할을 그대로 사용하기에는 문제점이 있다. 그것은분절의 경계에서 잘못된 분할을 할 염려가 있다는 것이다. 이러한 현상은 특히 짧은 무성음과 긴 유성음이 있을 때 무성음과 유성음의 경계가 유성음쪽으로 움직이게 된다. 따라서 이와 같은 현상을 방지해야 한다. 이를 위해 본 논문에서는 분절의 중간점을 기준점으로 하는 1회의 시간제한 집단화(timeconstrained clustering)를 행하도록 하는 방법을 채택하였다. 시간제한 집단화라고 하는 것은 집단화를 했을 때 각 집단이 시간상에서 서로 섞여 있어서는 안된다는 조건을 만족시키는 집단화 과정이다. 즉 다음의 조건을 만족하도록 분절의 경계를 구하는 것이다. 구해진 분절의 경계를 $\{b_1, b_2, \dots, b_i, \dots, b_M\}$ 이라고 할 때 i 가 j 보다 클 때 $b_j < b_i$ 를 만족하면서 분절들은 다음의 왜곡을 최소화 한다.

$$\text{distortion} = \sum_{i=0}^{M-1} \sum_{n=b_i+1}^{b_{i+1}} d(A_n, X_i) \quad (2)$$

단, x_i : 분절 i 의 중심

위의 조건을 만족시키기 위해 본 알고리즘에서는 집단화를 한 뒤 smoothing을 하도록 하였다(그림1, 2 참조). 이것은 집단화를 했을 때 시간제한을 만족하지 않는 경우가 있으므로 smoothing을 사용 이를 극복하도록 한 것이다.

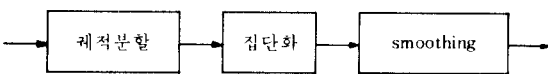


그림 1. 분할과정
Fig. 1. Segmentation process.

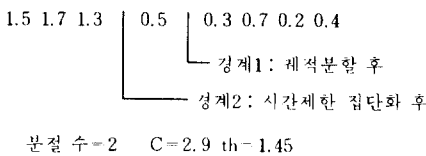


그림 2. 시간제한 집단화의 효과
Fig. 2. Effect of time constrained clustering.

3. 분절의 모델링

제적분할과 시간제한 집단화를 이용하여 찾아진 분절을 확률적으로 모델링하기 위해서는 집단에서 관측할 수 있는 값들과 이 값들의 확률 분포함수를 구해야만 한다. 가장 좋은 방법은 관측할 수 있는 값이 주어졌을 때 각 분절에서 정확한 확률분포 함수를 구하는 것이다. 일반적으로 Gaussian 분포로 모델링을 하게되는데 이것은 오차가 많고 정확한 분포를 구하기 위해서는 최적화 기법을 사용하는 등 어려운 점이 있다. 따라서 본 논문에서는 HMM과 같이 vector quantizer (V.Q)를 이용하여 관찰 값을 만들고 최대 유사도(maximum-likelihood)추정을 이용하여 각각 관측 값의 확률을 구하도록 하였다. 그런데 한가지 생각할 수 있는 것은 이미 분절이 나누어져 있음으로 codebook을 각 분절마다 만들 수도 있고 HMM과 같이 전체적으로 하나만 만들 수도 있다는 것이다. 그런데 알려진 실험결과에 의하면 대체로 codeword의 개수가 많을 수록 인식률이 향상되는 것으로 되어있다.¹⁰⁻¹² HMM과 같은 경우에는 codeword의 수가 늘어나면 계산량이 증가하는 단점이 있게 된다. 그러나 이미 찾아진 분절을 모델링함에 있어서는 각 분절을 따로 일정한 개수의 codeword를 만들어도 계산량은 전체적으로 하나만 만든 경우와 비교하여 그 계산량은 늘어나지 않으며 분절 개수 만큼 메모리의 양이 증가하게 된다. 그러나 패턴매칭의 경우에 비하여 이 메모리량의 증가는 큰 것이 아니며, 분절마다 codebook을 만든다면 패턴매칭의 방법과 비슷한 인식률을 얻을 수 있을 것이며 HMM과 비교하여 계산량을 줄이고도 더 높은 인식률을 얻을 수 있을 것이다. 따라서 분절마다 codebook을 만들기로 한다. 이러한 분할확률모델은 HMM에 비하여 제적분할의 부담이 있기는 하지만 codeword 개수에 대한 잇점과 학습과 인식시의 간편함을 생각하면 큰 부담은 없을 것으로 생각되며 대안어휘가 많을수록 그 영향은 적어질 것이다. 또한 단음절이 보다는 어음절의 경우 HMM보다 유리해 질 것이다. 만약 우리가 인식대상 어휘를 M개의 분절로 모델링을 한다면 학습단어의 모든 패턴을 분할한 후 각 분절에 대해 split-k-means 알고리즘(Linde-Buzo-Gray 알고리즘)¹⁵⁻¹⁶을 사용하여 N개의 codeword를 만든다. 즉

$$\begin{aligned} \text{codeword} : r(i, j) \quad & i = 1, 2, \dots, M, j = 1, 2, \dots, N(3) \\ M : & \text{분절의 개수} \\ N : & \text{codeword의 수} \end{aligned}$$

그다음 분할된 각 단어의 i 번째 분절을 $r(i, j)$ 에 encoding하여 관찰확률을 구하게 된다. 즉

$$P(i, j) = \frac{NC(r(i, j))}{\sum_{j=1}^N NC(r(i, j))} \quad (4)$$

NC(r(i, j)) : r(i, j)에 encoding된 갯수

으로 구하면 된다. 단어의 모델링은 그림 3 과 같다.

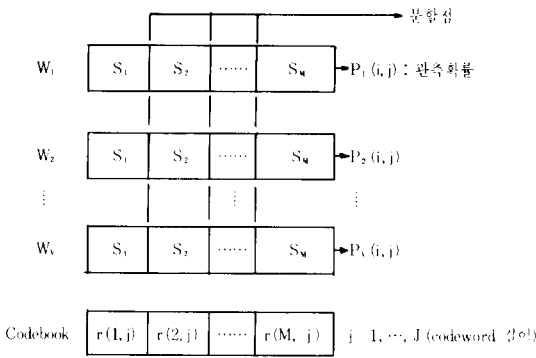


그림 3. 분할확률 모델에 의한 단어 모델
Fig. 3. Segment probability model.

3. 분할확률모델의 인식과정

분할확률 모델에 의한 인식 시스템은 그림 4 와 같다. 즉 음성신호가 들어오면 끝점 검출 뒤 LPC 분석을 통해 parcor 계수를 구한뒤 이 특성 파라미터를 이용하여 M개의 분절을 구하게 된다. 각 구해진 분절의 원소를 그 분절에 해당하는 codebook에 V, Q를 한뒤 encoding된 codeword index에 해당하는 확률을 곱해감으로써 특정 단어에 대한 확률을 구하게 된다.

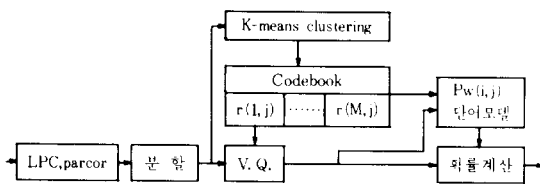


그림 4. 분할확률 모델인식 시스템
Fig. 4. Recognition system using segment probability model.

이때 확률 계산과정은 아래와 같다. 입력된 음성신호의 분할된 분절원소의 V, Q한 관측열은 아래와 같이 나타낼 수 있다.

$$O : \overset{s_1}{O_{11}O_{12}\dots O_{1n_1}}, \overset{s_2}{O_{21}\dots O_{2n_2}}, \dots, \overset{s_M}{O_{M1}\dots O_{Mn_M}} \quad (5)$$

단, n_i : 분절 S_i 의 길이

그러면 이 관측된 열로부터 각 단어에 대한 확률 P(O|W)을 구하면 된다. 그런데 분할된 각 분절을 어떻게 취급하느냐에 따라 두가지의 방법을 생각할 수 있다.

방법1 : 각 분절의 길이를 고려하지 않는 방법
이 방법은 음성인식에서 일반적으로 쓰이는 방법과 같다. 즉 각 분절의 길이에 따른 정규화를 하지 않고 V, Q index에 의한 확률을 계속 곱해 나간다. 단어 W에 확률 P_1 은

$$P_1 = \prod_{i=1}^M \prod_{j=1}^{n_i} P_w(i, j) \quad (6)$$

이다.

방법 2 : 각 분절의 길이에 따른 영향을 고려한 경우

이 방법은 고립단어의 인식에서 중요한 것은 어느 state의 길고 짧음에 있는 것이 아니라 어떤 확률 값으로 각 분절을 지나갔느냐에 있다는 사실에 기인한 것이다. 따라서 한 분절에서의 확률은 그 분절에 속하는 원소들의 평균확률로 표시하여 다음과 같다. 즉 확률 P_2 는

$$P_2 = \prod_{i=1}^M \left(\frac{1}{n_i} \sum_{j=1}^{n_i} P_w(i, j) \right) \quad (7)$$

이다.

4. 계산량의 비교

패턴정합 방법과 HMM 그리고 분할확률모델을 사용한 경우의 인식시 계산량을 간단히 비교하여 보자. 디지털 시스템에서 계산속도를 좌우하는 것은 덧셈보다는 곱셈이라고 할 수 있으므로 곱셈의 양만을 비교하기로 하자. 패턴정합의 방법은 주로 DTW에 의한 계산량이 대부분이며 HMM에 있어 계산량은 V, Q의 encoding 과정과 Viterbi 알고리즘에 의한 것이다. 분할 확률 모델에 있어서는 분할과 확률 계산량이 주요 계산요소이다. 따라서 위의 관계를 고려하여 곱셈의 양을 구하여 보면 아래와 같다.

- i) DTW의 경우 : $C = Q \cdot V \cdot N \cdot 2 \cdot W \cdot P$
 - ii) HMM의 경우 : $C = M \cdot N \cdot P + V \cdot N \cdot 3 \cdot S$
 - iii) 분할확률모델의 경우 : $C = M \cdot N \cdot P + N \cdot P \cdot 2 + V \cdot N$
- 단, Q : template 수
V : 인식 단어의 갯수
N : 프레임 수

- W : DTW의 window 폭
- P : LPC 차수
- M : codeword의 수
- S : 상태의 갯수

위의 경우에서 HMM과 분할에 의한 인식모델은 학습할 때 로그값을 취한 확률을 저장하여 덧셈으로 바꾸어 그 계산량을 더욱 줄일 수 있다. 위의 식들을 고려하여 볼 때 제안한 모델이 패턴정합 방법보다 그 계산량을 많이 줄일 수 있으며 HMM과는 비슷함을 알 수 있다. 만약 인식대상 어휘와 각 어휘당의 기준패턴 수가 늘어난다면 패턴정합 방법에 비한 계산시간의 감소율은 더 커질 것이다.

III. 실험 및 결과

1. 실험 데이터의 채집

본 논문에서는 위에서 제안한 모델의 타당성을 검토하기 위하여 한국어 숫자음을 대상으로한 화자독립 인식실험을 수행하였다. 음성의 채집은 남성50명의 화자가 10개의 숫자음을 각각 2번씩 발음하여 모두 1000개의 발음을 대상으로 하여 실험하였다. 음의 분석은 4KHz 6 차의 Butterworth 저역여파기를 통과한 후 10KHz 14bits A/D 변환을 하였다. A/D 변환을 한 신호는 200 샘플 단위의 프레임으로, 각 프레임으로, 각 프레임은 100 샘플씩 중첩하여 나누었다. 인식 파라미터는 10차의 LPC 분석을 하여 parcor 계수를 인식에 사용하였다.

2. 각 인식모델의 실험결과 및 검토

표 1, 표 2, 표 3은 각 패턴정합, HMM 그리고 제안한 분할확률 모델을 이용한 경우의 인식 결과를 보여주고 있다. 각 표에서 보여주는 인식의 결과는 50명의 화자가 2번씩 발음한 1000개의 단어중 한번의 것은 학습시에 나머지 500개는 인식시에 사용하여 얻어진 것이다. 즉 세 경우에 대해 같은 화자 그리고 같은 학습, 인식 데이터를 사용 실험하였다. DTW를 이용한 패턴정합의 실험에 있어서는 Sakoe와 Chiba가 제안한 대칭 가중합수를 갖는 warping 함수를 사용하였으며 창(window)의 폭은 6프레임으로 하였고 집단화의 과정으로서는 초기화하지 않는 K-means iteration의 방법¹⁷⁾을 이용하였다. 그리고HMM의 경우는 left-to-right 모델¹⁸⁾을 사용하였으며 Baum-Welsh 알고리즘과 Viterbi 알고리즘을 이용하여 학습 및 인식실험을 하였다. 표 2와 표 3은 상태와 분절의 갯수가 3, 4, 5, codebook당 codeword의 갯수가 16, 32, 64인 경우의 hidden Markov model과 분할확률 모델의 인식률을 보여주고 있다. 이 표에 의하면 두 알고리

표 1. 패턴정합에 의한 인식률

Table 1. Recognition rate of pattern matching method.

기준패턴수	1	2	3	4	5
인식률(%)	94.8	95.0	97.2	97.6	96.8

표 2. HMM에 의한 인식률

Table 2. Recognition of HMM.

상태 수	3			4			5		
codeword수	16	32	64	16	32	64	16	32	64
인식률(%)	90.4	89.4	92.0	90.0	93.6	91.4	91.0	90.0	91.6

표 3. 분할 확률 모델에 의한 인식률

Table 3. Recognition rate of segment-probability model.

분 절 수	3			4			5			
codeword 수	16	32	64	16	32	64	16	32	64	
인식률 (%)	방법1	93.6	95.6	96.0	94.2	94.4	95.6	93.2	96.8	92.4
	방법2	94.0	96.0	97.0	94.8	95.0	96.6	92.6	97.8	95.0

들은 같은 계산량을 가지면서도 분할확률 모델이 HMM보다 2~7% 정도 인식률이 높은 것을 알 수 있다. 이와같이 분할확률 모델이 hidden Markov model에 비하여 계산량을 늘리지 않고도 인식률이 향상된 것은 각 분절마다 codebook를 만들 수 있다는 잇점에 기인하는 것으로 생각되며 제안한 모델이 HMM보다 수학적으로 더 좋다는 것을 의미하는 것은 아닐 것으로 생각된다. 그리고 표 2에 의하면 각 분절의 길이에 의해서 확률을 정규화 했을 때가 이를 고려하지 않았을 때보다 대체로 0.4~1% 정도 향상된 것을 알 수가 있다. 이것은 고립단어의 인식에 있어서 중요한 것은 각 음소의 길이에 있는 것이 아니라 어떠한 음소를 가지고 있는가가 중요하다는 것을 보여주고 있다. 특히 짧은 무성음 구간을 갖고 있는 "구", "칠", "사"와 같은 경우에 인식률이 향상되어 시간 정규화가 필요함을 확인할 수 있었다.

한편, 패턴정합을 이용한 인식실험에서는 기준패턴의 갯수가 4개, window의 길이가 6일때 97.6%의 인식률을 얻을 수 있었는데, 분할 확률모델의 경우는 분절의 갯수가 5, codeword 갯수가 32일 때 97.8%로서 DTW를 이용한 패턴 인식방법에 필적할 만 한

결과를 보이고 있었다. 그리고 이때의 계산량은 대략 1/7정도로 줄어들음 III-4절의 결과를 이용하여 추산 할 수 있었다. 이 계산량의 감소비는 인식 단어 수가 늘어나고 기준패턴의 수가 많아질수록 더욱 커질 것으로 생각되어진다. 따라서 제안한 방법이 패턴정합에 비하여 그 계산량을 줄이면서도 비슷한 인식률을 보여 패턴정합 방법을 대체할 만하다고 생각되며 또한 고립단어의 인식에 있어서는 HMM과 같은 계산량을 가지고도 더 나은 인식 결과를 얻을 수 있었다.

IV. 결 론

본 논문에서는 패턴정합과 HMM의 단점과 고립단어의 특수한 성격을 고려하여 간단한 모델로써 높은 인식률을 얻을 수 있는 분할화물 모델을 제시하고 인식실험을 통하여 그 성능을 비교 검토하였다. 분할화물 모델은 결정적으로 상태를 전이시키는 방법으로서 미리 나누어진 분절을 벡터 양자화기를 이용 모델링하는 방법이다. 결정적으로 상태를 전이시키기 위해 케적분할을 사용하였으며 케적분할의 단점인 경계효과를 보강하기 위하여 시간제한 집단화를 이용하였다. 실험결과 비슷한 계산량을 갖을 때 HMM보다 높은 인식률을 얻을 수 있었으며 패턴정합에 의한 방법과 비슷한 인식률을 얻을 수가 있었다. 따라서 제안한 방법이 그 계산량이나 인식률에 있어서 HMM과 패턴매칭 방법을 대체할 수 있으리라 생각된다. 그리고 실험대상이 된 한국어 숫자음과 같은 단음절어 보다는 여러 음절로 된 단어의 경우 그 성능은 HMM과 패턴매칭에 비해 더욱 좋으리라고 생각된다. 그러나 분할을 위해 본 논문에서 사용한 방법은 개선의 여지가 있으며, 더 좋은 알고리즘을 개발하는 것이 차후의 과제로 남아 있다.

參 考 文 獻

[1] J.K. Baker, "The Dragon system-An overview," *IEEE Trans. on ASSP*, vol. ASSP-23, no. 1, pp. 24-29, Feb. 1975.

[2] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. on ASSP*, vol. ASSP-26, no. 1, pp. 43-49, Feb. 1978.

[3] L.R. Rabiner and A.E. Rosenberg, "Considerations in dynamic time warping algorithms for discrete word recognition," *IEEE Trans. on ASSP*, vol. ASSP-26, no. 6, pp. 575-582, Dec. 1978.

[4] C. Myers and L.R. Rabiner, "Performance tradeoffs in dynamic time warping algorithm for isolated word recognition," *IEEE Trans. on ASSP*, vol. ASSP-28, no. 6, pp. 623-635, Dec. 1980.

[5] S. Haltsonen, "Improved time warping method for discrete utterance recognition," *IEEE Trans. on ASSP*, vol. ASSP-33, no. 2, pp. 449-460, Apr. 1985.

[6] L.R. Rabiner and S.E. Rosenberg, "Speaker independent recognition of isolated word using clustering techniques," *IEEE Trans. on ASSP*, vol. ASSP-27, no. 4, pp. 336-349, Aug. 1978.

[7] S.E. Levinson and L.R. Rabiner, "Interactive clustering techniques for selecting speaker-independent reference templates for isolated word recognition," *IEEE Trans. on ASSP*, vol. ASSP-27, no. 2, pp. 134-141, Apr. 1979.

[8] L.R. Rabiner and J.G. Wilson, "Consideration in applying clustering techniques to speaker independent word recognition," *J.A.S.A.*, vol. 66, pp. 623-637, Sep. 1979.

[9] J.G. Wilpon and L.R. Rabiner, "A modified K-means clustering algorithm for use in isolated word recognition," *IEEE Trans. on ASSP*, vol. ASSP-33, no. 3, pp. 587-594, June 1985.

[10] L.R. Rabiner and B. Huang, "An introduction to hidden Markov models," *IEEE ASSP Magazine*, pp. 4-16, Jan. 1986.

[11] L.R. Rabiner and S.E. Levinson, "On the application to vector quantizer and hidden models to speaker independent isolated word recognition," *B.S.T.J.* vol. 62, no. 4, pp. 1075-1105, Apr. 1983.

[12] S.E. Levinson and L.R. Rabiner, "An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition," *B.S.T.J.* vol. 62, no. 4, pp. 1035-1073, Apr. 1983.

[13] G. David, "The viterbi algorithm," *IEEE Proc.* vol. 61, no. 3, March 1973.

[14] M.H. Kuhn and H.H. Tomaszewski, "Improvement in isolated word recognition," *IEEE Trans. on ASSP*, vol. ASSP-31, pp. 157-167, Feb. 1983.

[15] Y. Linde, A. Buzo and R.M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. on COM*, Jan. 1980.

[16] B. Juang and D.Y. Wonbg and .A.H. Gray, "Distortion performance of vector quantization for LPC voice coding," *IEEE Trans. on ASSP* , vol. ASSP-30, no. 2, pp. 294-304, Apr. 1982.

[17] 김진영, 성평모, "초기화하지 않는 K-means Iteration을 이용한 한국어 고립단어 인식," 한국전기전자공학회 학술대회 논문집, pp. 7-10, 1988. *

— 著 者 紹 介 —



金 泰 永 (准會員)

1962年 4月 26日生. 1986年 2月 서울대학교 전자공학과 졸업 공학사학위 취득. 1988年 2月 서울대학교 대학원 전자공학과 졸업 공학석사학위 취득. 현재 서울대학교 전자공학과 박사과정 재학중.

주관심분야는 디지털신호처리임.



成 宏 模 (正會員)

1947年 2月 1日生. 1973年 5月 서독 아헨(Aachen)공대 Vordiplom 1977年 10月 서독 아헨공대 Dipl-Ing. 1982年 9月 서독 아헨공대 Dr-Ing. 공학박사학위 취득. 1983年 7月~현재 서울대학교 전자공학과

교조교수, 부교수. 주관심분야는 초음파공학, 약기음합성, 음성신호처리 등임.