

DMS 모델을 이용한 한국어 음성 인식

(Korean Speech Recognition using Dynamic Multisection Model)

安 泰 玉*, 邊 龍 圭*, 金 淳 協*

(Tae Ock Ann, Yong Kyu Byun, and Soon Hyob Kim)

要 約

본 연구는 DMS(dynamic multisection) 모델을 이용한 불특정 화자에 의한 한국어 음성 인식으로써, 연속된 시간 상에서 발음된 단어 패턴 중 비슷한 특징을 대표하는 특징 벡터와 시간 정보를 이용해 DMS 모델을 만드는데, 시간 정보를 얻기 위해 backtracking 기법을 사용하는 모델링 알고리즘을 제안한다.

모델의 각 상태는 시계열로 표현되는데 각 상태는 시간 정보와 상태를 대표하는 특징 벡터를 가지고 있으며 각 단어의 패턴에서 대표 특징 벡터는 어떤 상태에서 패턴간의 거리가 최소가 되는 특징 벡터로 정한다.

인식 대상어는 전국 146개 DDD 지역명을 선정하였으며, 특징 파라메타로는 12차 LPC Cepstrum 계수를 사용하였고, 모델의 상태는 8로 하였으며, 시간정보를 이용하는데 가중치는 0.2를 사용하였다.

실험 결과, DMS 방법에 의한 인식률은 94.8%이고, 이는 비교 실험한 MSVQ(multisection vector quantization) 방법에 의한 인식률 89.3%보다 향상되었음을 보여주고 있다.

Abstract

In this paper, we proposed an algorithm which used backtracking method to get time information, and it be modelled DMS (Dynamic Multisection) by feature vectors and time information which are represented to similiar feature in word patterns spoken during continuous time domain, for Korean Speech recognition by independent speaker using DMS.

Each state of model is represented time sequence, and have time information and feature vector. Typical feature vector is determined as the feature vector of each state to minimize the distance between word patterns.

DDD Area names are selected as recognition wocabulary and 12th LPC cepstrum coefficients are used as the feature parameter. State of model is made 8 multisection and is used 0.2 as weight for time information.

Through the experiment result, recognition rate by DMS model is 94.8% , and it is shown that this is better than recognition rate (89.3%) by MSVQ (Multisection Vector Quantization) method.

I. 서 론

음성 인식에 있어서 소·중규모(20~500)의 단독어에 있어서는 음소나 음절 단위와 같이 subword 단위로 패턴 매칭을 행하는 것보다 단어 단위로 음성 인

*正會員, 光云大學校 電子計算氣工學科
(Dept. of Computer Eng., Kwangwon Univ.)
接受日字: 1990年 8月 14日

식을 행하는 것이 더 실용적이라 할 수 있다. 따라서, 본 연구에서는 146개의 DDD 지역명을 인식 대상으로한 단어 단위의 한국어 단독음 인식에 관한 연구를 행하려 한다.

단어 단위의 인식 방법에는 DP(dynamic programming) 매칭 방법¹¹⁾이나 HMM(hidden markov model) 모델^{12,13)}을 이용한 확률적인 방법등이 있는데 DP 매칭 방법은 인식시간 및 표준패턴을 위해 많은 데이터량이 필요하며, HMM에 의한 확률적인 방법의 경우에는 높은 확률값을 가진 모델로 만들기 위해 학습에 필요한 데이터 및 학습하는데 많은시간이 걸린다는 단점이 있다.

단어 단위의 패턴 매칭에서는 파열음이나 파찰음등이 짧게 발음되는 음소인데 이들 음소는 단어 전체 구간에 걸쳐 누적 거리를 계산하면 거리계산에 기여하는 정도가 작게 된다. 따라서, 음소계열의 잡음이 하나만 다른 단어쌍의 경우에 있어서는 그런 단어 사이에 인식이 상당히 어렵다. 그러나, 이러한 경우는 DDD 지역명 인식시에 정주, 청주, 성주 등 몇 지역을 제외하고는 거의 없으며, DTW나 HMM을 이용한 인식의 경우에도 마찬가지로 인식에 어려움이 있다.

본 연구에서는 단어 패턴의 연속된 대표값(시간정보와 대표 특징 벡터)의 정보만 사용해서 음성 인식을 행함으로써 파열음이나 파찰음 등과 같은 미소한 특징 벡터의 변동이 무시되지만 거기에는 시간정보가 포함되어 있으므로 다른 인식 방법에 비해 인식 성능이 떨어지지 않으며, 단어 단위의 인식 방법인 DP 매칭 방법에 비해 데이터량 및 인식시간을 줄이고, HMM에 의한 확률적인 방법에 비해 학습시간, 학습데이터 및 인식시 데이터량을 줄이는 방법으로 backtracking에 의해 시간 정보를 구하는 DMS 모델을 이용한 음성 인식 모델을 제안한다.

이 모델은 DP 매칭이 각 프레임에 따른 시계열의 특징 벡터에 의해 표준 패턴이 형성되는데 비해 시계열의 대표값에 의한 소수의 상태로 표현되는 모델이라는 점에서 특히 level building DTW(dynamic time warping)^{15,16)}과 비슷하며, 또한 level building DTW에서의 backtracking 기법을 이용하여 시간정보를 추출한다. 여기에서, 소수의 상태의 시계열로 표현시에는 DP 매칭이 시간에 따른 각 프레임의 특징 벡터와의 매칭으로 인해 인식률을 높여 주는데 비해 몇 프레임에 대한 대표값으로 대치됨으로 시간 정보와 특징 벡터를 함께 이용함으로써 인식률을 향상시키는 모델이다.

이 모델은 개념적으로 frame의 길이가 다른 몇 개

의 section으로 나누어 clustering 기법^{17,18)}을 이용하여 그 대표 특징 벡터를 구하며, 또한 시간정보를 이용하는 방법이므로 VQ codebook^{9,10)}의 열로써 시간 정보를 포함하는 MSVQ codebook^{11,12)}에 의한 인식과 비교한다.

II. 모델 작성

모델을 작성하기에 앞서 먼저 <제주>에 대해 예를 들어보자. <제주>을 발음했을 때에 시간의 변화에 따른 음소의 변화를 4상태로 나누어서 생각할 수 있다. <제주>라는 단어의 각 음소를 각각 한 상태에 대비시켜 보자. 이 때 상태 1에는 <ㅈ>음소에 해당하는 몇 개의 frame이 속하게 될 것인데 이 특징 벡터를 대표할 수 있는 1개의 특징 벡터만을 구하고 싶을 때는 clustering 방법에 의해 1개의 특징 벡터를 구할 수 있으며, 또한 이 상태에 속하는 frame 수를 전체 frame 수로 나누어 <제주>라는 단어에서 차지하는 비율을 시간 정보로 사용할 수 있다. 또한 상태 2에서도 마찬가지로 <ㅊ>에 해당하는 대표 특징 벡터와 시간정보를 구할 수 있으며, 상태 3에서도 다른 음소로써 <ㅊ>, 상태 4에서 <ㅌ>에 대한 대표 특징 벡터 및 시간 정보를 구할 수 있다.

위의 예에서와 같은 방법으로 각 단어마다 몇개의 동적인 section으로 나누어 대표 특징 벡터와 시간 정보를 가지는 모델을 만들 수 있는데 이를 DMS 모델이라 한다.

이와 같은 방법으로 만든 모델이 MSVQ에 의해 section을 나누는 것보다 합리적인 대표 벡터 추출방법이 된다는 것을 알 수 있다. 본 연구에서 이와 같은 DMS model을 이용한 인식 시스템을 제안한다.

1. 패턴 매칭 알고리즘

각각의 단어 모델 M을 작성하는데 있어서, 단어마다 J개 상태의 시계열로 표현되며, 상태 j($1 \leq j \leq J$)는 그 상태를 대표하는 특징 vector R_j 와 시간정보 P_j 가 등록되어 있다.

여기에서, 미지의 입력 시험 패턴 T_i ($1 \leq i \leq I$, I는 frame 수)와 각 단어 모델 M과의 거리 $DIS(T_i, M)$ 를 아래와 같이 정의한다.

본 연구에서는 12차의 LPC cepstrum을 특징 벡터로 사용하였으며 시간정보는 단어 전체의 길이를 1로 하여 전체 길이 중 어떤 상태에 시험 패턴이 머무는 때까지의 비율이 얼마인가로 정의하여 사용하였다.

먼저 modelling시에 입력 음성 T를 j개의 section으

로 등간격으로 분할한 다음 j번째 section을 단어 model의 j번째 상태에 대응시킨다.

이 때 시험 pattern이 j번째 상태에서 다음 상태로 천이시의 종료 frame을 e_j 로 표시한다.

따라서 $e(0)=0$ 이고 $e(J)=I$ 이다.

이 때 DIS(T, M)은

$$DIS(T, M) = \min_{e(0), \dots, e(J)} \left\{ \sum_{j=1}^J [S(j) + P(j)] \right\} \quad (1)$$

로 정의된다.

여기에서 S(j)는 대응하는 입력 음성의 j section의 특징 벡터와 단어 모델의 j번째 상태의 특징 벡터 간의 거리이고,

$$S(j) = \sum_{i=m_j^{j-1}+1}^{m_j^j} d_v(t_i, m_j) \quad (2)$$

이다. 여기에서 함수 $d_v(a, b)$ 는 2개의 벡터 a, b간의 거리로 표시된다. 또한 여기에서 p(j)는 j번째 section까지의 frame수를 전체 frame 수로 나눈 비율값이다.

따라서

$$P(j) = W * d_s(p(j), i) \quad (3)$$

으로 정의한다. 여기에서

$$d_s(p(j), i) = |p(j) * I - i| \quad (4)$$

로 주어진다. 또한 W는 가중치로써 본 연구에서 몇 가지 실험을 거쳐 0.2로 주었다.

이상의 식으로 부터 DIS(T, M)은 각 section의 대표 특징 벡터에 의한 거리 D와 시간정보인 P에 의한 거리에 의해 평가되어짐을 알 수 있다.

그런데 식(1)에 의하면 최적의 경로를 찾기위해 전체를 다 평가해 본다는 것은 계산량이 방대하게 됨으로 DTW에서 처럼 점화식을 이용하여 소구간 경로 세약을 돕으로써 DIS(T, M)를 계산한다. 따라서 DIS(T, M)을 알고리즘에 이용하면

$$D(i, j) = d_v(t_i, m_j) + \min \begin{cases} D(i-1, j) \\ D(i-1, j-1) + P(j-1) \end{cases} \quad (1 < i \leq I, 1 < j \leq J) \quad (5)$$

이다. 즉,

$$DIS(T, M) = D(I, J) \quad (6)$$

로 계산된다. 여기서 D(i, j)는 입력 음성 T부분의 pattern $t_1, t_2, \dots, t_i (1 \leq i \leq I)$ 를 j(1 ≤ j ≤ J)개의 section으로 분할해, 단어 model의 i frame-j 상태와 대응시켜서 매칭했을때의 누적 거리를 나타낸다.

본 연구에서 제안하는 DMS model에서는 식(5), (6)

의 누적 거리 D에 시간정보에 관한 거리 P를 포함시키는데

$$D(i, j) = d_v(t_i, m_j) + \min \begin{cases} D(i-1, j) \\ D(i-1, j-1) + P(j-1) \end{cases} \quad (1 < i \leq I, 1 < j \leq J) \quad (7)$$

로 해서, DIS(T, M)를 구하는 것이다.

식(7)은 입력 음성의 j-1 section까지의 frame수에 대해 전체 길이에 대한 비율이 시간정보에 의한 비율에 가장 가까울 때 가장 거리값이 적어지므로 j 상태로 천이할 확률이 높아진다. 결국 시간 정보에 관한 거리 p에 관해서는 식(3)처럼 구한다.

(기호)

i: 입력 음성 T의 frame 번호 (1 ≤ i ≤ I)

j: 단어 model M의 상태 번호 (1 ≤ j ≤ J)

$d_v(t_i, m_j)$: t의 I frame의 특징 벡터 t_i 와 m의 J 상태의 특징 벡터 m_j 와의 국부적인 거리

$d_s(p(j), i)$: j상태의 마지막 frame 수와 실제 입력 패턴의 수와의 차에 절대값을 취한다.

D(i, j): i frame과 j상태까지의 누적거리

W: 계속 시간의 거리에 대한 가중치

DIS(T, M): 입력 음성 T와 단어 model M의 거리

$e(j)$: 입력 음성의 j section의 종료 frame으로, 알고리즘을 계산하고 최종적으로 backtracking에 의해서 결정된다.

2. 단어 Model 작성방법

단어 model은 단어 model 작성용으로 발생된 음성(훈련 data)을 가장 효율 좋게 표현하게끔 작성할 필요가 있다. 훈련 data로 사용되는 단어 패턴을 $T_1, T_2, T_3, \dots, T_i, \dots, T_I$ 이라고 하고, 1절에서 나타낸 점화식에서 계산한 단어 패턴 T_1 와 단어 model M과의 거리를 DIS(T_1, M)이라고 한다. 이 때

$$D_{min} = \sum DIS(T_i, M) \quad (8)$$

를 평가 함수로 해서, 이 D_{min} 이 최소로 되게끔 단어 model M를 작성한다. 그러기 위해서는 다음 두 가지를 고려할 필요가 있다.

1) 단어 패턴을 분할해서 각 상태에 할당하는 방법(그림 1)

2) 할당된 각 상태의 패턴에서 대표하는 특징벡터를 작성하는 방법

먼저 2)에 대해서는 단어 패턴 $T_1, T_2, T_3, \dots, T_i, \dots, T_I$ 의 j상태에 할당된 frame의 특징 벡터의 centroid를 계산해 이것을 단어 model의 j상태를 대표하는 특징 벡터로 한다.

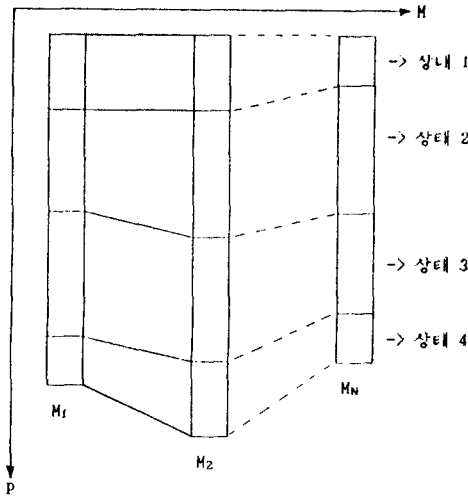


그림 1. 훈련용 데이터의 상태 할당의 예
 Fig. 1. Example of state assignment of training data.

그림 2의 흐름도와 같은 방법으로 단어 model을 최적화 하였다.

먼저 초기 단어 model은 훈련 data를 시간 축상에서 등분할하여, 각 상태에 할당된 frame의 특징 벡터의 centroid를 계산해서 구했다. 이때 계산 시간은 어느 상태에도 같은 것으로 되어있다.

예를 들면, 매칭 경로가 그림 3 같이 되었다고 하자.

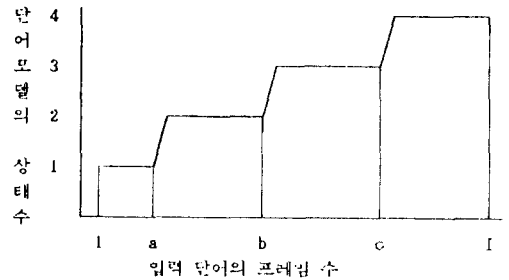


그림 3. 매칭 경로의 예 (J=4)
 Fig. 3. Example of matching path (J=4).

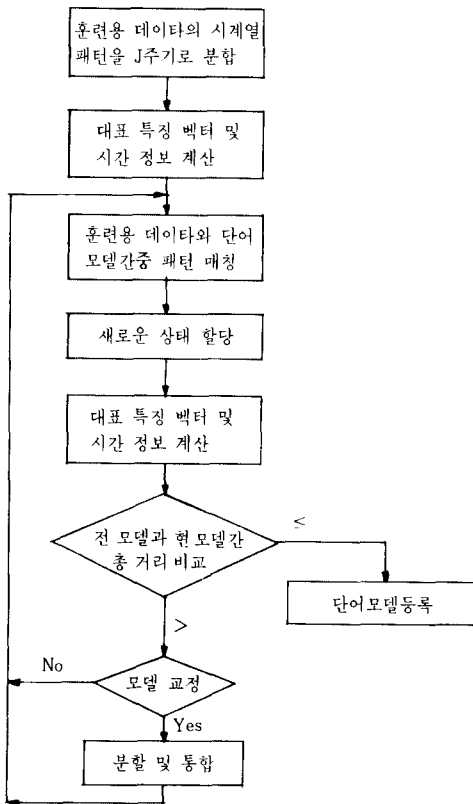


그림 2. 단어 모델 작성 방법
 Fig. 2. Method of word model making.

즉, $\alpha(0)=0, \alpha(1)=a, \alpha(2)=b, \dots, \alpha(J)=l$ 로 되었을 때의 1상태는 1에서 a frame까지, 2상태는 a+1에서 b frame까지 등등으로 할당된다.

그런 다음 새롭게 할당된 frame의 특징 벡터에서 centroid를 계산해 단어 model을 갱신한다. 동시에 각 상태에 할당된 frame 수를 전 frame 수로 나눈 비율을 단어 model의 시간 정보로 해서 등록한다.

이것으로 단어 model을 한번 갱신한 것이 된다. 이하는 패턴 매칭에 의한 단어 model 갱신이 수렴할때까지 반복한다.

일반적으로 이와 같은 반복 계산을 계속하는 것만으로 경계의 국소적인 변경만의 문제를 해결하려고 한다면 수렴후의 해(단어 model)는 최적해의 보장은 없고, Local minimum에 빠져버릴 가능성이 있다.

거리서 경계를 크게 움직여 보다 양호한 해를 구하기 위해서는, 수렴후의 단어 model을 M_x 로 했을때, 그림 2의 분할 통합부에서는 아래에 기술한 처리를 실시해 최적해에 가깝게 하고 있다.

1) 상태내의 거리(distortion) (각 훈련 패턴 T_i 와 단어 model M_x 간의 $S(j)+T(j)$ 의 합)가 최대가 되는 상태 j_{max} 를 검출해, 이 상태를 두 개의 상태 j_a 와 j_b 로 등분할하고, 각각의 centroid를 계산한다.

2) 연속하는 두 개의 상태의 특징 벡터 사이의 거리가 최소가 되는 조 (j_c, j_{c+1})를 검출해서 한 개의 상태로 통합하고, 다시 centroid를 계산한다.

3) 1), 2)의 처리에 의해 작성된 단어 모델을 M_y 이라 했을때, $DIS(T_i, M_x) > DIS(T_i, M_y)$ 이라면 M_y 는 적어도 M_x 보다 최적해에 가까운 해가 된다. 따라서 M_y 을 이용해 앞에 기술한 반복 계산을 수행한다.

역으로 $DIS(T_i, M_x) < DIS(T_i, M_y)$ 이면 M_x 이 훨씬 최적해에 가깝다고 판단됨으로 M_x 을 등록시킨다.

III. 실험 결과 및 고찰

DMS 모델에 의한 불특정 화자의 음성 인식을 하는데 있어서, 대상 어휘로는 146개의 DDD 지역명을 선정하였으며, 3명의 화자가 각각 3번씩 발음한 것 중 각각 2번씩 발음한 것으로 DMS 모델의 대표 특징 벡터를 구했으며, 각 화자의 훈련 데이터로 사용하지 않은 나머지 데이터로 인식 실험을 행했다.

1. 인식 시스템 구성

본 논문에서 제안하는 DMS 모델에 의한 음성 인식 시스템을 그림 4에 나타내었는데, 실험에 사용된 모든 DDD 지역명 데이터는 마이크에 의해 입력된 음성 신호로써, 샘플링 주파수를 8KHz로 하였으며, 3.5KHz의 저역 여파기를 통과한 후 12 bit A/D 변환기를 거쳐 음성 데이터를 구한 다음, 시작점과 끝점을 검출한 후, 특징 파라미터로 LPC 계수에 의한 LPC cepstrum 계수를 구해 각 단어의 모델을 구한 다음 이 모델을 기초로 인식 실험을 행했다.

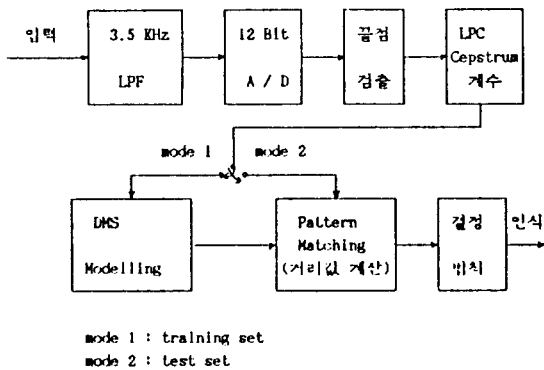


그림 4. 인식 시스템의 구성도
Fig. 4. Block diagram of recognition system.

2. 계산식 및 가중치 결정

특징 벡터간의 거리인 d_v 은 자승한 거리를 사용하였으며, 시간 정보에 의한 거리값 계산식 및 이 때 이용되는 가중치 w 는 다음과 같은 방법에 의해 결정되었다. 각각의 단어에 대한 모델을 DMS 모델링 방법에 의해 만든 후 인식시에도 DMS 모델링시와 같은 방법으로 각 section의 대표 특징 벡터를 구한 후 각 상태마다 선형으로 거리값을 계산하여 인식률이 가장 높은 가중치를 택했으며, 또한 계산식은 식(4)에 의한 계산 방법과 다음에 표시할 식(16)과 비교하여 이 중 가장 인식률이 좋은 식(4)에 의한 계산식과 가중치 0.2를 택했다.

$$d_s(p(j), i) = |\log(p(j) * I/i)| \tag{9}$$

이것이 각 상태간의 선형 비교시 인식률이 가장 좋았다는 것은 가장 적당하게 상태를 형성해 주는 식이며, 가중치라 할 수 있어서 이것을 택했다. 이 실험에 대한 인식률은 다음 표1, 표2에서 보여준다.

표 1. 식(16)에 대한 가중치 변화에 의한 인식률
Table 1. Recognition rate by change of weight for exp.(9).

가중치	10	20	30	40	50	60	70	80	90
오류/총수	7/27	5/27	3/27	1/27	0/27	1/27	1/27	1/27	2/27

(a) 인식 단어를 9개 선정의 경우
(3명의 화자가 1번씩 발음)

가중치	45	50	55
오류/총수	12/57	12/57	12/57

(b) 인식 단어를 19개 선정의 경우
(3명의 화자가 1번씩 발음)

표 2. 식(4)에 대한 가중치 변화에 의한 인식률
인식 단어를 19개 선정의 경우
(3명의 화자가 1번씩 발음)

Table 2. Recognition rate by change of weight for exp.(4).

가중치	0.1	0.2	0.3
오류/총수	17/57	10/57	13/57

3. 인식 실험 결과

인식 실험시 사용된 데이터는 MSVQ에 의한 방법이나 DMS 모델에 의한 방법이나 똑같이 마이크로

입력된 데이터로 3명의 화자가 전국 DDD 지역명 146개 단어를 각각 3번씩 발음한 총 1314개로 그 중 각 화자가 두번씩 각 단어에 대해 발음한 876개는 단어의 모델을 작성하는데 이용되었으며, 나머지 각 화자가 1번씩 발음한 438개는 인식하는데 이용하였다.

실험에서는 상태를 8로 하여 시험하였으며, 기존의 MSVQ에 의한 방법과 비교하였다. 본 실험에서 비교되는 기존의 MSVQ codebook 작성에 의한 인식에서는 각 section의 codeword가 두개씩으로 총 16개의 특징 벡터를 표준 벡터로 하여 인식하는데 반해 본 연구에서는 각 상태에 하나의 대표 특징 벡터와 1개의 시간 정보(스칼라 값)만으로 구성되어 기억용량이 줄어든다.

본 연구에서 제안하는 DMS 모델을 이용한 인식에서는 분할·통합부를 둔상태와 분할·통합부를 안둔 상태로 나누어서 인식실험을 행했다. 인식의 결과는 표 3에 나타내었다.

표 3. 인식 결과
Table 3. Recognition results.

인식방법 화자	MSVQ에 의한 인식률	분할·통합부 없는 DMS 모델에 의한 인식률	분할·통합부 있는 DMS 모델에 의한 인식률
화자 1	95.0%	99.0%	98.0%
화자 2	87.0%	92.5%	93.0%
화자 3	86.0%	92.5%	92.5%

4. 고찰

표 1에서 보는 바와 같이 시간 정보에 대한 식을 식(9)을 이용하여 인식시켜본 결과가 표 2를 나타내는 식(4)을 사용한 인식 방법에 비해 인식률이 낮다. 그 이유는 식(9)의 경우 입력 패턴이 상태에 머물 시간 정보와의 비가 1에 가까우면 거리값이 0이 나오고 또한 그 비율이 커질 경우에도 그 결과는 그리 큰 변화를 주지 못한다. 그러나 식(4)에 의한 방법에 의하면 그 비가 1일때 시간정보에 의한 거리값이 0 이고 그 앞 뒤 프레임들은 선형적으로 거리값이 증가한다. 따라서, 이러한 이유에 근거하여 식(4)을 이용하여 시간 정보를 주는 것이 더 타당한 것 같으며, 가중치에 있어서는 식(4)를 사용하는 경우에 0.2를 중심으로 0.1과 0.3에서 떨어짐을 알 수 있다.

또한 표 3에 의하면, MSVQ codebook을 이용한 인

식보다 본 연구에서 제안하는 DMS 모델을 이용한 패턴 매칭 방법이 인식률을 높여 주고 있으며, 분할·통합부를 둔 경우의 인식률과 이런 부분을 두지 않은 경우의 비교에 있어서는 분할·통합부를 두는 경우의 유효성을 발견할 수 없다. 실제 비교에 있어서 분할·통합부를 둔 경우에 오차가 하나 더 많았다.

IV. 결 론

본 연구에서는 DMS 모델에 의한 불특정 화자의 음성인식을 하였다. 인식 대상으로는 146개 DDD 지역명을 선택하였으며, MSVQ codebook에 의한 인식 실험과 비교하였다. 본 연구에서 제안하는 DMS 모델은 단어 패턴을 몇 개의 section으로 나눈 후 그 section을 대표할 수 있는 대국적인 정보만을 추출하는데 이 때 clustering 기법을 이용하여 대표 특징 벡터를 구하고, Level building DTW에서와 마찬가지로 backtracking에 의해 시간 정보를 구하는 모델이다.

이 DMS 모델에 의해 구해진 대표 특징 벡터와 시간 정보는 기존의 DTW나 level building DTW 에서 특징 벡터간의 거리만 계산한 것에 비해, 입력 패턴이 들어올 때 시간 정보까지를 이용하게 된다. 따라서, 이 시간 정보를 이용하게 됨으로 DTW와 같은 패턴 매칭 기법에 비해 몇개의 대표 특징 벡터만으로 인식이 가능하다.

이 때 시간 정보를 거리값으로 사용하는데 가중치는 실험에 의해 0.2가 가장 좋았으며, 분할·통합부를 둔 경우와 안둔 경우로 나누어 실험한 결과는 분할·통합부를 두지 않은 경우가 인식된 단어가 하나 더 많은 것으로 나타났는데 이것에 이하면 분할·통합부를 두는 경우의 유효성을 발견할 수 없었다.

본 실험에서는 상태가 8인 경우의 DMS와 section을 8로 나누고 각 codebook마다 2개씩의 codeword를 가져 총 16개의 codeword로 이루어진 MSVQ codebook을 이용한 인식 방법과 비교하였는데, 비교결과 MSVQ codebook에 의한 인식률이 약 89.3%인데 비해 훨씬 개선된 94.8%의 인식률을 나타내었다. 따라서, 시간 정보를 고려한 가변 길이의 특징 벡터를 추출해서 인식시키는 방법이 좋은 인식 방법임을 알 수 있었다.

參 考 文 獻

[1] 김순협, "한국어 음성의 분석과 자동 인식에 관한 연구" 박사 논문, 연세대학교 대학원, 1982. 12.

- [2] L.R. Rabiner, B.H. Juang, "An Introduction to Hidden Markov Models," *IEEE ASSP MAGAZINE* Jan. 1986.
- [3] L.R. Rabiner, S.E. Levinson, M.M. Sondhi, "On the Application of Vector Quantization and Hidden Markov Models to Speaker-independent, Isolated Word Recognition," *Bell System Technical Journal*, vol. 62, no. 4, April 1983.
- [4] Kai-Fu Lee, *Automatic Speech Recognition: The Development of the Sphinx System*, Kluwer Academic Publishers, 1989.
- [5] C.S. Myers, R.R. Rabiner, "Connected Digit Recognition Using a Level-Building DTW Algorithm," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-29, pp. 351-363, June 1981.
- [6] C.S. Myers, R.R. Rabiner, "A Level Building Dynamic Time Warping Algorithm for Connected Word Recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 284-297, Apr. 1981.
- [7] J.T. Tou, R.C. Gonzalez, *Pattern recognition Principles*, Addison-Wesley Publishing Company, Inc, 1974.
- [8] S.E. Levinson, L.R. Rabiner, A.E. Rosenberg and J.E. Wilpon, "Interactive Clustering Techniques for Selecting Speaker-Independent Reference Techniques for Isolated Word Recognition," *IEEE Trans. on ASSP* vol. 27, no. 2, pp. 134-141, Apr. 1979.
- [9] R.M. Gray, "Vector Quantization," *IEEE ASSP Magazine*, vol. 1, pp. 4-29, Apr. 1984.
- [10] Y. Linde, A. Buzo, and R.M. Gray "An algorithm of Vector Quantizer Design," *IEEE Trans. Comm.*, vol. COM-28 pp. 84-95, Jan 1980.
- [11] D.K. Burton, J.E. Shore, J.T. Buck, "Isolated-Word Speech Recognition Using Multisection Vector Quantization Codebooks," *IEEE Trans. of Acoustics, Speech, Signal Processing* vol. ASSP-33, no. 4, August 1985.
- [12] 이성권, "VQ를 이용한 DDD 지역명 인식에 관한 연구," 석사학위 논문, 광운대학교 대학원, 1989. 12.

 著 者 紹 介

安 泰 玉 (正會員) 第27卷 第7號 參照
 현재 광운대학교 전자계산기
 공학과 박사과정

邊 龍 圭 (正會員) 第27卷 第9號 參照
 현재 서울산업대학 전산과
 교수



金 淳 協 (正會員) 第27卷 第7號 參照
 현재 광운대학교 전자계산기
 공학과 교수