한국어 단독 숫자음 인식을 위한 DTW 알고리즘의 비교

(Comparison of the Dynamic Time Warping Algorithm for Spoken Korean Isolated Digits Recognition)

* 홍 진 우 (Hong, Jin Woo)
** 김 순 협 (Kim, Soon Hyob)

Abstract

This paper analysis the Dynamic Time Warping (DTW) algorithms for time normalization of speech pattern and discusses the Dynamic Programming (DP) algorithm for spoken Korean isolated digits recognition.

In the DP matching, feature vectors of the reference and test pattern are consisted of first three formant frequencies extracted by power spectrum density estimation algorithm of the ARMA model.

The major differences in the various DTW algorithms include the global path constraints, the local continuity constraints on the path, and the distance weighting/normalization used to give the overall minimum distance. The performance criterias to evaluate these DP algorithms are memory requirement, speed of implementation, and recognition accuracy.

요 약

이 연구는 음성 패턴의 시간 정규화를 위해 사용된 다양한 DTW 알고리즘들을 설명하고, 한국어 단독 숫자음 인식에 적합한 DP알고리즘을 논하는데 있다.

DP정합에 필요한 표준 패턴과 시험 패턴의 특징 벡터들은 ARMA 모형의 PSD 추정 알고리즘을 이용하여 추출한 제1, 제2, 제3포르만트 주파수이다.

다양한 DTW 알고리즘들은 전체 최소 거리를 찾기 위해 사용된 전체 경로 제약과 소구간 연속성 경로 제약 그리고, 웨이팅 함수 및 정규화에 의해 구별된다. 이들 DP알고리즘의 상이점을 비교하기 위한 인식 수행 기준은 기억 용량, 수행 시간, 인식률로 결정하였다.

* 한국전기통신연구소 (KETRI) 연구원
** 광운대학 전자계산기공학과 교수

1. 서 론

인간과 기계와의 대화를 위한 음성의 자동 인식 과정에서 인간이 발성하는 음성은 개개인의 발성 속도와 발성 지속 시간의 차이로 인하여 음성 패턴에서 시간축의 비선형적 변동을 야기시킨다.

그러므로, 이러한 변동의 제거 및 시간축의 정규화는 단독 음성 인식 연구에 중요한 과제로 제시되었고, 이에따라 다양한 DTW 알고리즘(Dynamic Time Warping algorithm)을 이용한 DP(Dynamic Programming)법이 연구되어지고, 널리 이용되어 거의 완벽한 인식 수행을[1] 얻게 되었다.

그러나 한국어 음성에서 DP알고리즘을 이용한 인식 수행은 별로 없고, DP알고리즘을 이용한 극소수의 인식 수행마저도 다양한 DTW알고리즘에 수반되는 각종 파라미터(Parameter) 및 요인(factor)들을 전혀 고려하지 않고, 임의로 선택하여 수행하고 있기때문에 인식률의 감소를 가져올 뿐만아니라 수행 시간을 단축시키지 못하고 있다.

따라서, 본 연구의 목적은 한국어 음성 인식을 위한 다양한 DTW알고리즘의 상이점을 설명하고, 한국어 단독 숫자음을 인식대상으로 하여 각각의 DTW알고리즘을 이용한 DP법을 적용시키고, 그 파라미터 및 요인들을 비교 분석하는데 있다.

DP정합에 필요한 표준 패턴과 시험 패턴의 특징 벡터들은 ARMA모형(Autoregressive Moving Average model)의 PSD(Power Spectrum Density) 추정 알고리즘을 이용하여[3] 추출한 제1, 제2, 제3 포르만트 주파수(Formant frequency)이다.

2. DP인식 알고리즘

2-1. DTW알고리즘의 원리

음성은 적당한 특징 추출에 의한 특징 벡터들의 계열(sequence)로써 표현[5]된다.

표준 패턴(Reference pattern) $\vec{R}(m)$과 시험 패턴(Test pattern) $\vec{T}(n)$의 특징 벡터는 다음 식으로 표현되는 다중 벡터[2]이다.

$$\vec{R}(m) = [R_1, R_2, \cdots, R_j, \cdots, R_M]$$
$$\vec{T}(n) = [T_1, T_2, \cdots, T_i, \cdots, T_N] \quad (1)$$

본 연구에서는 선형 예측법에 의한 ARMA 모형 방식으로 부터 추출한 3개의 포르만트 주파수(f_1, f_2, f_3)를 시험 패턴과 표준 패턴의 특징 벡터(T_i 또는 R_j)로 사용하였다.

(1)식에서 패턴 T(n)과 R(m)은 각각 n축과 m축에 따라 변화하고, 음성 패턴들은 같은 영역에 존재하며 그것들 사이의 특징 벡터 차이는 점 C(k)의 계열로서 다음 식과같이 표현될 수 있다.

$$F = C(1),\ C(2),\ \cdots, C(k),\ \cdots, C(K) \quad (2)$$

이 계열 F를 워핑 함수(warping function)라 하며 그림 1과 같이 표현된다.(1차원의 예)

따라서 DTW문제는 워핑 함수를 이용하여 최소화 거리가 되는 최적 경로

$$m = W(n) \quad (3)$$

을 찾는 것이다. 식(3)에서 n과 m사이의 관계를 간단한 함수적 표현으로 만들기 위해 공동 시간축 k를 도입하고, n, m을 k의 함수로써 표현하면 다음 식과 같다.

$$n = i(k), \quad k = 1, 2, \cdots, K$$
$$m = j(k), \quad k = 1, 2, \cdots, K \quad (4)$$

여기서 K는 공동 시간축의 길이이고 그림 1

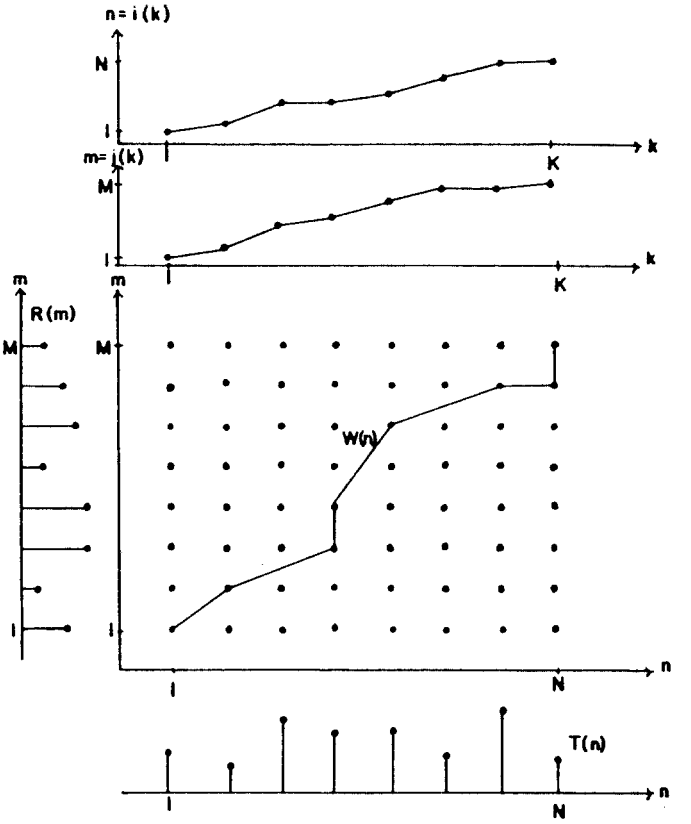

그림 1. 경로 m=W (n) 을 따른 T (n)과 R (m)의 warping경로 및 공동시간축 k

Fig. 1. An example of a warping T(n) to R(m) and a function of common time axis k via the path m=w(n).

의 상단에 $i(k)$와 $j(k)$가 k의 함수로써 표현되는 것을 보인다.

식(4)의 변수적 표현을 근거로 (n, m)평면에서 최적 경로를 얻기 위한 워핑 함수는 다음과 같은 DTW알고리즘의 제약 조건들을 만족하여야 한다.

1. 끝점 제약

표준 패턴과 시험 패턴의 끝점이 정확히 결정되었다면 워핑 함수의 시간축 경계점 제약은 다음과 같다.

$$\begin{aligned} &i(1)=1,\ j(1)=1 \text{ ; 시작점 제약} \\ &i(K)=N,\ j(K)=M \text{ ; 끝점 제약} \end{aligned} \tag{5}$$

2. 소구간 연속성 제약

최적 경로를 결정하기 위한 방법으로 시간축의 과다한 압축이나 팽창을 방지하기 위해 다음과 같은 소구간 제약이 수반된다.

(가) 단조 증가 조건

$$\begin{aligned} i(k+1) &\geq i(k) \\ j(k+1) &\geq j(k) \end{aligned} \tag{6}$$

(나) 소구간 경로의 기울기에 대한 제약 조건

워핑 함수의 소구간 경로 기울기는 여러가지 형태로 표현 가능하고, 이로 인하여 다양한 DTW알고리즘이 제시된다.

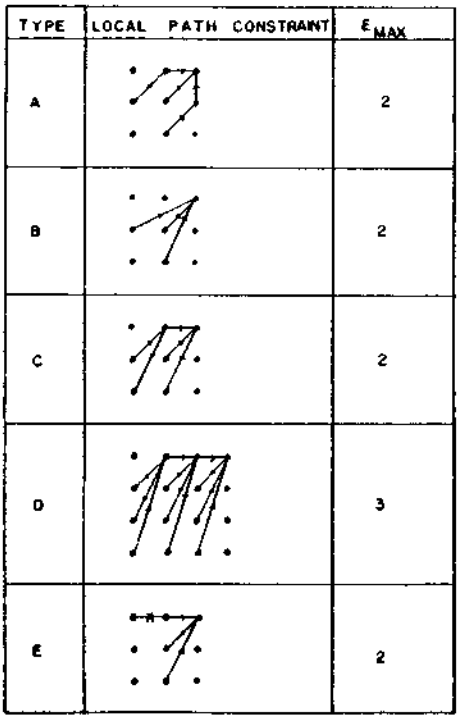

그림 2. 소구간 경로 제약의 5가지형

Fig. 2. The variations of the local path constraint.

본 연구에서는 그림 2와 같은 경로의 기울기를 선정하여 실험하였다. 이 그림에서 최대 기울기 E_{max} 는 다음 식으로 표현된다.

$$E_{max} = \max\left[\sum_{l=1}^{L} \beta_l / \sum_{l=1}^{L} \alpha_l\right] \quad (7)$$

여기서 L은 소구간 경로에서 출발점과 목적점을 잇는 선분의 수이고, α와 β는 각각 소구간 경로가 갖는 시험 패턴과 표준 패턴의 프레임 수이다.

D형은 A형, B형, C형의 응용이며 E형은 1프레임이상 같은 면을 지나지 못하도록 경로 제한(X표시 부분)을 한 형으로써 다음과 같은 제약 조건을[7] 갖는다.

$$i(k+1) - i(k) = 1$$

$$j(k+1) - j(k) = \begin{cases} 0,1,2, \\ if\ j(k) \neq j(k-1) \\ 1,2, \\ if\ j(k) = j(k-1) \end{cases} \quad (8)$$

3. 전체 경로 제약

소구간 경로 제약 때문에 최적 워핑 경로가 진행할 수 있는 (n, m)평면이 제한되는데 이 때문에 일체화의 잘못을 최소로 줄이고, 계산에 필요한 기억 용량의 감소와 수행 시간의 절약을 꾀할 수 있다.

본 연구에 적용시킨 전체 경로 제약에 대한 허용 범위를 수식으로 표현하면 식(9)와 같다.

$$1 + \frac{1}{E_{max}}(i(k) - 1) \leq j(k) \leq 1 + E_{max}(i(k) - 1)$$

$$M + E_{max}(i(k) - N) \leq j(k) \leq M + \frac{1}{E_{max}}(i(k) - N) \quad (9)$$

4. 축 변환

워핑 함수를 구할 때 $n = i(k), m = j(k)$로 변수를 정하는 경우와 $n = j(k), m = i(k)$로 변수를 정하는 경우가 있다. 본 연구에서는 여러 참고 문헌을[1) 2) 3)] 조사한 결과 전자의 경우가 후자의 경우보다 더 좋은 인식률을 보이고 있어 전자의 경우를 택하여 인식수행을 실험하였다.

2-2. 거리 측정법

시험 패턴과 표준 패턴사이의 두 특징 벡터 차이는 두 벡터들 사이의 거리 측정(dis-

tance measurement)으로 얻어질 수 있는데 거리 d(c)는

$$d(c)=d(i(k),\ j(k))=\|\vec{T}_i-\vec{R}_j\| \quad (10)$$

로 표시되며 최적 워핑 경로를 얻기 위해 사용되는 전체 거리 측정은 다음 식과 같이 정의 된다.

$$D(i(k),\ j(k))=\frac{\sum_{k=1}^{K} d(i(k),\ j(k))\cdot W(k)}{N(w)} \quad (11)$$

여기서 $D(i(k),\ j(k))$는 경로의 길이 K에 따른 전체 거리로 주어진 함수이고, $d(i(k),\ j(k))$는 시험 패턴의 프레임 $i(k)$와표준 패턴의 프레임 $j(k)$의 소구간 거리이며, $W(k)$는 k번째 소구간 경로의 웨이팅 함수이고, $N(w)$는 웨이팅 함수 W의 함수인 정규화 요소(normalization factor)이다. 최적 경로의 정의는 전체 거리 $D(i(k),\ j(k))$를 최소로 하는 경우로써 윗식으로 부터직접 구해질 수 있다.

$$D_T=\min_{(K,\ i(k),\ j(k))}[D(i(k),\ j(k))] \quad (12)$$

윗 식들을 인식 실험에 적용하기 위해서는 3가지 함수 즉, 소구간 거리 함수 d, 웨이팅 함수 W 그리고,정규화 요소 N(w)가 선정되야만 한다. 본 연구의 인식 실험에 적용한 함수의 선정은 다음과 같다.

1. 소구간 거리 함수의 선정

본 연구에서는 특징 벡터로 제1,제2,제3 포르만트 주파수를 추출하였기때문에 Euclidean거리측정방법을 선정하여 사용하였다. Euclidean거리 측정[5]은

$$d=\left[\sum_{p=1}^{3}(f_{i,p}^{(T)}-f_{j,p}^{(R)})^2\right]^{\frac{1}{2}} \quad (13)$$

로 표현되며 여기서 $f_{i,p}^{(T)}$는 시험 패턴의 특징 벡터이고, $f_{j,p}^{(R)}$은 표준 패턴의 특징 벡터이다.

2. 웨이팅(Weighting)함수의 선정

웨이팅 함수는 단지 소구간 경로에만 의존하며 다음과 같은 4가지 형태가 있다.

$$\begin{aligned}
&(\text{I})\ W(k)=\min(i(k)-i(k-1),\ j(k)-j(k-1))\\
&(\text{II})\ W(k)=\max(i(k)-i(k-1),\ j(k)-j(k-1))\\
&(\text{III})\ W(k)=i(k)-i(k-1)\\
&(\text{IV})\ W(k)=i(k)-i(k-1)+j(k)-j(k-1)
\end{aligned} \quad (14)$$

여기서 $i(0)$와 $j(0)$는 0이라 가정한다.

본 연구에서는 4가지 웨이팅 함수를 C형의 소구간 경로 제약에 각각 적용시켜 인식 실험을 하여 비교하였다. 이때 이상의 방법으로 웨이트를 설정할 때 그림3의 (a)와 같이 O(zero)의 웨이트를 갖는 경우가 생기게 되는데 이때의 소구간 경로는 전체 거리에 전혀 영향을 주지않아 최소화 거리 측정에 위배된다.[2] 이러한 현상을 없애기 위하여 smoothed웨이팅 함수를 적용시켜 그림3의 (b)와 같이 웨이트를 설정하였다.

3. 정규화 요소의 선정

정규화 요소 N(w)의 선정은 전체 거리가 소구간 평균 거리로 이루어 진다는 제약에

의해 결정된다. 또, 이것은 표준 패턴과 시험 패턴의 길이 모두에 독립적이다. 이런이 유때문에 정규화는 식(15)와 같이 정의된다.

$$N(w)=\sum_{k=1}^{K} W(k) \tag{15}$$

본 연구에서는 정규화 요소의 순환 계산이 어려운 점을 감안하여 또, 수행의 효율성을 높히기 위해 4가지 경우의 웨이팅 함수에대한 정규화 요소 N(w)를 모두 N으로 선정하였다.

2-3. DTW알고리즘을 이용한 DP의의 수행

소구간 최적 경로에 따라 점 (n,m)에 이르는 누적된 거리 함수는 다음 식과 같다.

$$D_A(n,m)=\min_{(i(k),j(k),k)}[D_A(n',m')+d(n',m'),(n,m)\cdot W(k)] \tag{16}$$

여기서 $n'\leq n$, $m'\leq m$이며, d는 (n', m')에서 (n,m)에 이르는 거리이다.

한편, 정규화된 최소 경로 거리 D_T는

$$D_T=\min_{(i(k),j(k),k)}\frac{\sum_{k=1}^{K} d(i(k),j(k))\cdot W(k)}{N(w)} \tag{17}$$

이고, N(w)가 경로에 독립적이기 때문에 거리는 다음 식과 같이 쓸 수 있다.

$$D_T=\frac{\min_{(i(k),j(k),k)}\left[\sum_{k=1}^{K} d(i(k),j(k))\cdot W(k)\right]}{N(w)} \tag{18}$$

그러므로,

$$D_T=\frac{D_A(N,M)}{N(w)} \tag{19}$$

로 표현할 수 있다.

3. 인식실험 및 결과

3-1 인식 실험

본 연구의 인식 실험에 사용된 숫자음(/0/ ~/9/)은 성인 남성 3인에 의해 발성된 음성으로써 발성 시간을 0.3 ~ 0.6 (sec) 정도로 제한하였다.

인식 대상 어휘 수는 3인이 각각 한숫자음에 대해 매회 5번씩 반복 발성하여 5회에 걸쳐 녹음한 것중 임의로 선정한 150개 숫자음이다.

인식 실험에 대한 전체 시스템은 그림 4와 같다.

한편, 다양한 DTW알고리즘의 수행을 비교하기 위하여 본 연구에서는 다음과 같이 표현되는 비교 기준을 설정하였다.

1. 기억 용량

누적된 거리 함수 $D_A(n,m)$을 계산하기 위해서 저장되어야 하는 벡터들의 수와 크기

2. 수행 시간

최적 경로를 계산할 때 DP알고리즘에 의해 인식에 소요되는 시간

3. 인식률

최종 인식 결정에서 시험되어진 단독 단어수와 인식된 단어수의 비

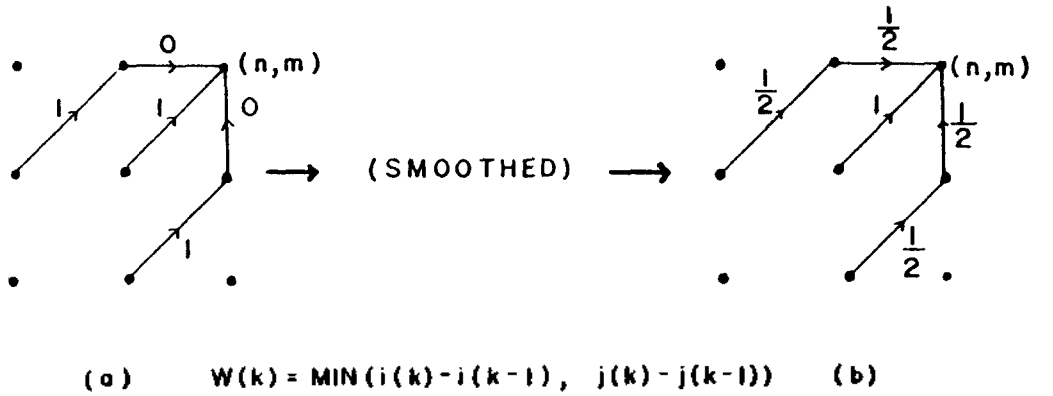

그림 3. A형 소구간 경로 제약의 smoothed weighting 함수

Fig. 3. Example of the smoothed weighting function on the path of type A constraint.

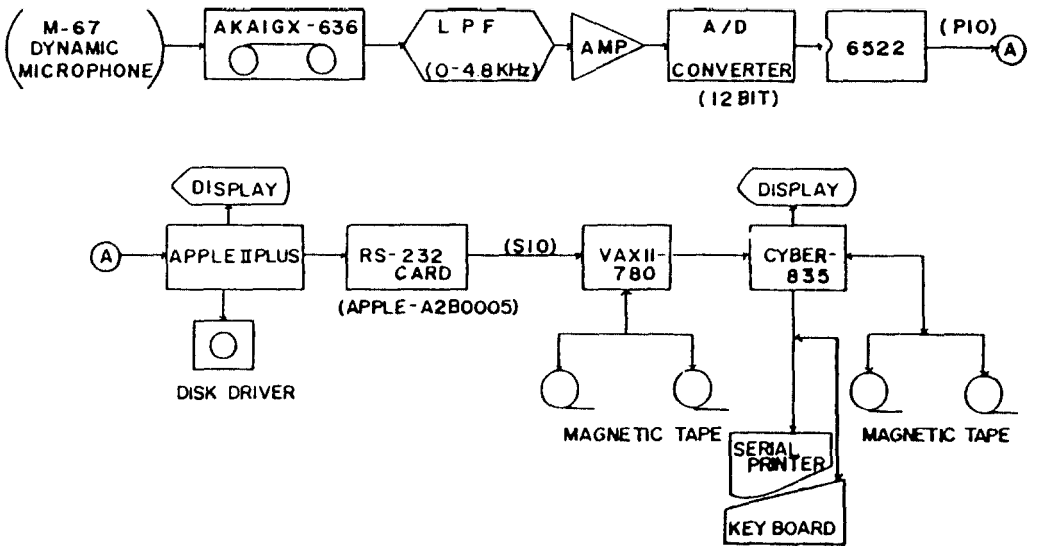

그림 4. 한국어 음성 분석 및 인식에 관한 시스템

Fig. 4. A system for Korean speech analysis and recognition.

3-2 결과 및 고찰

1. 기억 용량

기억 용량의 관점에 대한 DTW알고리즘들 간의 차이는 누적된 거리, 소구간거리,그리고 특별한 제약 조건을 주게되는 사이드(side) 정보에 필요로 하는 벡터들의 수와 크기에 관련된다. 표 1에 각각의 소구간 경로에 대해 필요로 하는 벡터의 수를 나타냈다.

표로부터 필요로 하는 전체 벡터 수가 A형과 C형이 3개, B형과 E형이 2개, D형이 5개임을 알 수 있으며, E형은 소구간

TYPE / VECTOR NO.	A	B	C	D	E
ACCUMULATED DISTANCE	2	2	2	3	1
LOCAL DISTANCE	1	0	1	2	0
SIDE INFORMATION	0	0	0	0	1
TOTAL	3	2	3	5	2

표 1. 기억 용량에 필요로 하는 벡터의 수

Table 1. The number of vectors for memory requirement.

경로가 프레임과 프레임 사이를 연속해서진행할 때 같은 패턴의 직선 평면을 연속으로 진행하여 수행할 수 없도록 제약을 주는 1개의 사이드 정보가 필요함을 알 수 있다.

2. 수행 시간

한 단어를 인식시키는데 필요한 전체 수행 시간을 전체 경로 제약을 준 경우와 주지 않은 경우에 대해 고찰하였다. 실험 결과로 산출된 각각의 소구간 제약형의 인식 수행 시간은 표2와 같다.

이 표로 부터 전체 경로 제약을 준 경우가 전체 경로 제약을 주지 않은 경우에 비하여 현저하게 인식 수행 시간을 단축시켰다.

CONSTRAINT / TYPE	GLOBAL PATH CONSTRAINT (SEC/WORD)	NO GLOBAL PATH CONSTRAINT (SEC/WORD)
A	2.094	2.521
B	1.938	2.133
C	2.163	2.504
D	2.825	3.683
E	1.977	2.306

표 2. 각각의 소구간 제약 형태에 대한 평균적 인식 수행 시간

Table 2. The overage time of the recognition for the local constraints.

3. 인식률

(가) 인식되어진 단어들의 끝점 프레임 길이를 계산하고, 이에 대한 시험 패턴과 표준 패턴사이의 평균적 거리 계산을 하여 그림 5와 같은 결과를 얻었다.

그림으로부터 프레임 길이의 차가 적으면 적을수록 평균적 전체 거리 계산없이 작아져서 오인식률(recognition error rate)이 감소하는 것과 가장 작은 평균적 전체 거리계산은 시험 패턴과 표준 패턴의 길이가 같을 때 얻어진다는 것을 알 수 있다.

(나) 4가지 형태의 웨이팅 함수를 C형의 소구간 제약 형태에 적용한 결과 표3과 같은 오인식률을 얻었다.

표로 부터 (Ⅲ)형의 웨이팅 함수를 DP법에 적용시켰을 때 가장 작은 오인식률이 얻어지며 (Ⅱ)형의 웨이팅 함수를 적용할 때 가장 큰 오인식률이 얻어짐을 알 수 있다. 따라서 본 연구에서는 각각의 소구간 제약 형태에 대한 오인식률을 계산하기 위해 (Ⅲ)형의 웨이팅 함수를 적용하였다.

(다) 전체 150개 단어를 시험 대상으로 하여 표4와 같은 오인식률을 얻었다.

전체 경로 제약을 주지 않은 경우의 오인

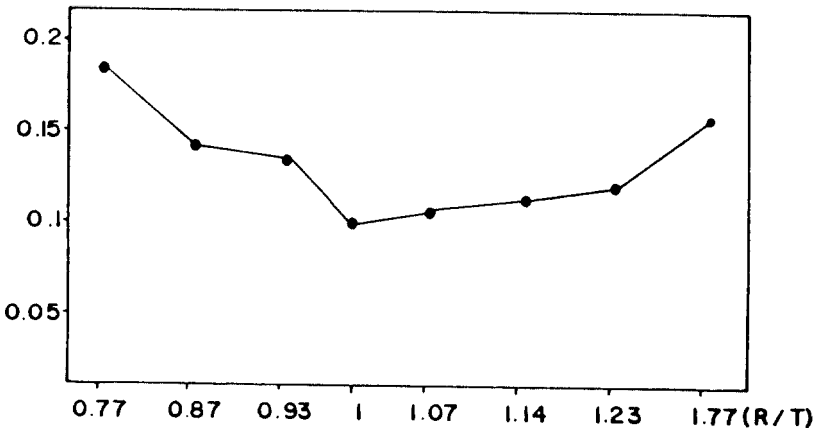

그림 5. 표준 패턴과 시험 패턴의 길이 비율에 따른 평균거리

Fig. 5. The effect of the average distance ratio between reference and test pattern.

WEIGHTING TYPE	ERROR RATE
Ⅰ	6 %
Ⅱ	8 %
Ⅲ	4 %
Ⅳ	5.3 %

표 3. 4가지 weighting 함수를 적용한 오인식률

Table 3. The recognition error rate for the 4 weighting function.

TYPE	GLOBAL PATH CONSTRAINT	NO GLOBAL PATH CONSTRAINT
A	6.7 %	18.6 %
B	6 %	7.3 %
C	4 %	9.3 %
D	4 %	14 %
E	2.7 %	4 %

표 4. 각각의 소구간 제약 형태에 대한 오인식률

Table 4. The recognition error rate for the local constraints.

식률은 전체 경로 제약을 준 경우의 오인식률보다 1.2배에서 3.5배까지 매우 높게 얻어지고 있으며, 전체 경로 제약을 준 경우 A형, C형, D형은 오인식률이 현저하게 감소하지만 B형과 E형은 약간의 감소를 보이고 있어 A형, C형, D형의 소구간 경로 제약에는 수행 시간의 단축 뿐만아니라 오인식률을 줄이기 위해 전체 경로 제약을 반드시 사용하여야 한다는 것을 고찰하였다.

4. 결 론

다양한 소구간 제약 형태와 웨이팅 함수 그리고, 전체 경로 제약을 적용하여 각각의 DTW알고리즘의 수행 시간과 오인식률을 비교 분석한 결과 다음의 결론을 얻었다.

1. 평행 사변형꼴의 전체 경로 제약을 DTW알고리즘에 적용시킴으로써 모든 소구간 경로 제약의 알고리즘에 대한 인식 수행 시간과 오인식률을 감소시킨다.

2. 전체 시험 단어에 대한 인식 수행의오인식률과 기억 용량에 필요한 벡터수는 E형이 가장 작다.

3. 4가지 웨이팅 함수중에서 (Ⅲ)형의웨이팅 함수가 모든 소구간 제약 형태에 적용되었을 때 가장 좋은 인식률을 얻었다.

4. 시험 패턴과 표준 패턴의 평균적 거리가 가장 작아져서 최적의 DTW알고리즘의수행을 얻을 수 있는 것은 두 패턴들 사이의프레임 길이 비가 1일 때이다.

Reference

1. Ihroaki Sakoe, and Seibi Chiba, "Dynamic Programming Algorithm Optimiza-

tion for Spoken Word Recognition," IEEE. Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-26, No. 1, pp. 43-49, Feb. 1978.

2. C. Myers, L.R. Rabiner and A.E. Rosenberg, "Performance Tradeoffs in Dynamic Time Warping Algorithms for Isolated Word Recognition," IEEE. Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-28, No. 6, pp. 623-634, Dec. 1980.
3. S.M. Kay, "A New ARMA Spectral Estimator," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-28, No. 5, pp. 585-588, Oct. 1980.
4. 김순협, "한국어 음성의 분석과 자동 인식에 관한 연구," 박사학위 논문, 연세대학교 대학원. 1982.12.
5. L.R. Rabiner and S.E. Levinson, "Isolated and Connected Word Recognition-Theory and Selected Applications," IEEE Trans. on Commications, Vol. COM-29, No. 5, pp. 621-639, May. 1981.
6. L.R. Rabiner and R.W. Schafer, Digital Processing of Speech Signal, Prentice-Hall, Inc., Englewood Cliffs, pp. 38-249, pp. 396-430, New Jersey, 1978.
7. M.K. Brown and L.R. Rabiner, "An Adaptive, Ordered, Graph Search Technique for Dynamic Time Warping for Isolated Word Recognition," IEEE Trans. On Acoustics, Speech, and Signal Processing, Vol. ASSP-30, No. 4, pp. 535-543, Aug. 1982.
8. L.R. Rabiner, A.E. Rosenberg, and S.E. Levinson, "Considerations in Dynamic Time Warping Algorithms for Discrete Word Recognition," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-26, No. 6, pp. 575-582, Dec. 1978.