

One stage MSVQ/DP를 이용한 음성 인식에 관한연구

A Study on Speech Recognition by One Stage MSVQ/DP

정 의 봉*

(Eui Bung Jeoung*)

요 약

본 논문은 One Stage MSVQ/DP를 제안하여 단어 인식을 수행하였다. 인식 대상 어휘로는 대학교 행정부서명 40개를 선정 하였고 인식을 위한 특징 파라메타로는 10차 LPC 캡스트럼 계수를 사용하였다. 본 연구에서 제안하는 One Stage MSVQ/DP 인식 시스템 이외에도 같은 데이터 상에서 LBDTW 인식 시스템, One Stage DP 인식시스템에 의한 음성인식 실험을 수행하였다. LBDTW와 One Stage DP알고리즘에 의한 인식율은 83.3%와 87.5%였으며 본 연구에서 제안한 MSVQ/DP에 의한 인식율은 91.6%였다.

abstract

This paper proposes One Stage MSVQ/DP method for word recognition system university administration branch names are selected for the recognition experiment and 10 LPC cepstrum coefficients is used as the feature parameter. Besides the speech recognition experiments by proposed method, for comparison with it, we perform the experiments on the same data by Level Building DTW and One Stage DP method.

The Recognition rates with the LBDTW and the One Stage method are 83.3% and 87.5%, but the recognition rate with the proposed method is 91.6%.

I. 서 론

음성인식을 실행하는 방법에는 음성에 포함된 언어적 특징에 대한 기준패턴을 미리 저장해 놓고 입력 음성과 비교하여 유사성을 판별하는 패턴 매칭법과 확률적인 모델링 방법 그리고 벡터 양자화 방법이 있

다. VQ는 기준패턴을 코드 북(code book)으로 작성하여 소규모의 벡터열을 저장하므로 데이터 양이 작고 인식시간이 빠른 일반적인 DP 매칭법과 비교해 인식율이 떨어지는 단점이 있었다. 이러한 여러가지 인식 알고리즘 중에서 가장 인식율이 좋은 방법의 하나인 다이나믹 프로그래밍 기술에 대해서 살펴보면, Vintysyuk [1]와 Rabiner [2]는 단독어 인식문제와 유사하게 최적화 문제로서 연결단어 문제를 형식화

*전북산업대학교 정보통신공학과
접수일자: 1993년 8월 13일

시켰다. 이러한 형식에서 가장 중요한 것은 적은 양의 음성정보와 음성학습이 용이하게 이루어지도록 기준패턴을 작성하는 것이다. 또 하나의 특징은 3가지 수행단계로 단어 경계 검출, 비선형적인 시간축 정렬, 그리고 인식이 동시에 수행되는 것이다. 연결 단어 인식을 살펴보면 Sakoe는 최적화 문제를 해결하기 위해 two-level 알고리즘을 독립적으로 유도해냈다. 첫번째 레벨에서 모든 기준패턴들은 패턴의 모든 분할부분에 대해 조직적으로 매칭 되는 것이다. 이러한 매칭단계는 단독어 인식에서의 비선형적 시간축 정렬과 같은 것이다. 두번째 레벨에서는 생성된 거리값을 가지고 미지의 단어열의 최적평가가 모든 단어열의 전체 거리값을 최소화 함으로써 구해진다. Myer와 Rabiner는 모든 단어열의 매칭이 연속적인 기준패턴들의 연결에 의해 수행될 수 있다는 점을 이용하였다. 이러한 알고리즘을 소위 Level Buidung DTW 알고리즘[2] 이라고 한다. 마지막으로 One-Stage 알고리즘[9]은 워핑경로를 하나의 지표로써 파라미터화 함으로써 워핑경로를 직접 최적화 범위의 한 함수로 취급하는데 기초한다. 그러나 연결단어 인식을 위한 이러한 DP 방법들도 전체적으로 인식율을 향상시킬 수는 있으나 여전히 비교횟수가 많아 계산시간이 길어지는 단점이 있다. 결국, 음성인식에 있어서는 인식율도 떨어 지지않고 기억용량도 적게 차지하는 알고리즘의 개발이 시급한 실정이다.

하지만 아직도 음성의 실제 데이터량이 많고, 음성의 시간변화율이 크며, 기준패턴의 최적화가 어려워 특징화자에만 국한되어 사용되는 실정이다. 본 연구에서는 이에 따라서 DP에 비해 인식율이 떨어지지 않고 VQ와 기억용량이 같으나 처리시간이 훨씬 줄어드는 MSVQ방법[5][7]의 시간정보를 이용해 기준패턴을 작성하였고 One Stage DP방법을 개선한 One Stage MSVQ/DP 알고리즘을 제안한다. 비교를 위하여 같은 데이터 같은 조건상에서 기존의 One Stage DP 패턴매칭 뿐만아니라 연결음성 인식방법인 LBDTW 패턴매칭과 함께 음성인식 실험을 수행하였다.

II. 전처리 및 음성 신호 분석

2.1 끝점검출

프레임 길이는 샘플링 주파수와 깊은 관계가 있다.

그 관계식은

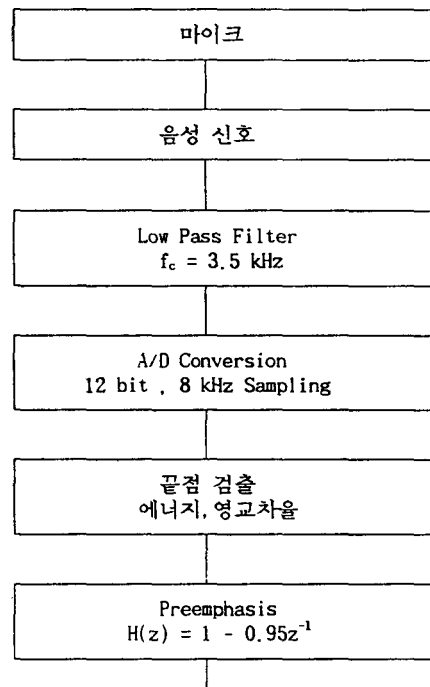
$$N = \delta \cdot F_s$$

로 나타낼 수 있다. 여기서 F_s 는 샘플링 주파수이고, δ 는 $10 \cdot 10^{-3} \leq \delta \leq 20 \cdot 10^{-3}$ 의 값을 갖는다. 본 연구에서는 프레임 길이 N 을 128샘플로 즉, 16ms 단위로 창함수를 사용하였으며 프레임 간격을 두지 않았다.

음성 인식의 경우, 한 단어의 음성이 묵음(silence)을 앞뒤로 하고, 그 사이에 발음된다는 것을 전제하기 때문에 단어 인식을 위해서는 묵음으로부터 음성 부분을 구별해 낼 필요가 있다. 본연구는 단시간 에너지와 단시간 영교차율을 이용하여 음성 구간을 검출한다.[2][9]

2.2 음성 특징 추출

끝점 검출 과정을 통해 음성구간을 추출한 후에는 각 프레임별로 음성의 특징을 추출하는데 본 연구에서는 LPC 계수보다 LPC 캡스트럼 계수[6]를 특징 벡터로 사용하는 것이 좋다는 고찰에 따라 특징 파라미터로 10차 LPC 캡스트럼 계수를 사용하였다. 음성 신호처리의 전체적인 흐름도를 그림 2.1에 나타내었다.



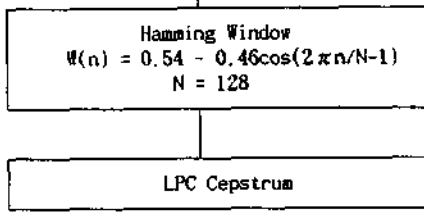


그림 2.1 음성 신호 처리의 전체 흐름도
Fig 2.1 Flow chart of speech signal analysis

III. LBDTW 인식 알고리즘

3.1 개요

LBDTW 알고리즘은 여러개의 단어를 연결하여 발음하므로써 구성된 입력패턴 중에서 한 단어가 종결되고 다른 단어가 시작될 가능성이 있는 점들을 표준패턴의 레벨을 하나씩 늘어가면서 탐색하여 그중 거리가 최소가 되는 단어열 외에도 거리가 최소에 가까운 단어들의 목록을 제공할 수 있다. L개의 단어로 이루어진 입력 문자열의 패턴을 $T(m)$, $m = 1, 2, \dots, M$ 이라고 하자. 여기서 $T(m)$ 은 m번째 프레임의 특징 벡터이고, M은 입력패턴의 총 프레임 수를 의미한다. 이 경우 Level Building DTW 알고리즘의 목표는 L개의 표준패턴의 열인 $R_{q(1)}(n), R_{q(2)}(n), \dots, R_{q(L)}(n)$ 을 찾는 것이다. 여기서 $q(k)$, $k = 1, 2, \dots, L$ 은 V개의 표준 패턴으로서 R_v , $v = 1, 2, \dots, V$ 중의 하나로써 $T(m)$ 과 super 표준패턴 $R^s = R_{q(1)}(n) \oplus R_{q(1)} \oplus R_{q(2)} \oplus \dots \oplus R_{q(L)}$ 과의 dynamic time warped 거리가 모든 $q(k)$ 에 대해 최소가 되도록 하는 값들이다. (\oplus 는 단어열이 연결됨을 의미한다). 보다 수식적으로 표현하여 $D_{q(1)q(2)\dots q(L)}(M)$ 이 입력패턴의 $T(m)$ 과 super 표준패턴 $R_{q(1)} \oplus R_{q(2)} \oplus \dots \oplus R_{q(L)}$ 사이의 누적된 DTW 거리라고 정의하면

$$D^* = \min_{q(1)q(2)\dots q(L)} [D_{q(1)q(2)\dots q(L)}(M)] \quad (3-1)$$

로 정의된 최소 거리값 D^* 를 구하는 작업으로 볼 수 있다.[2]

D^* 를 최소화하는 표준 패턴의 단어열의 지표(index) 들인 $q^*(1), q^*(2), \dots, q^*(L)$ 가 발음된 단어열과 가장 잘 match되는 단어열이 됨을 알 수 있다. 그림 3.2은 $L=1$ 레벨에서 입력패턴의 일부분과 가장 잘 match

되는 값을 찾기위해 V개의 표준패턴 R_v 가 어떻게 사용되는지를 보여준다.

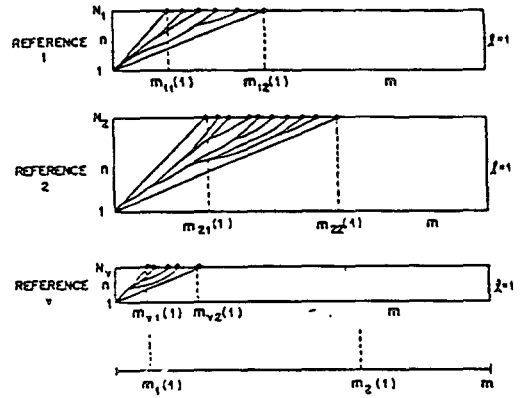


그림 3.1 LBDTW 알고리즘의 첫번째 레벨에서의 매칭 영역과 끝영역의 예

Fig 3.1 The match regions and ending regions used in the first level of the LBDTW.

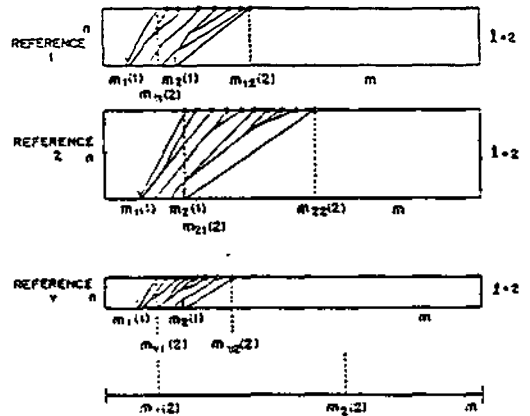


그림 3.2 LBDTW 알고리즘의 두번째 레벨에서의 매칭 영역과 끝영역의 예

Fig 3.2 The match regions and ending regions used in the secode level of the LBDTW.

그림 3.1에서 보는바와 같이, 표준패턴 R_1 에 대해서 이 알고리즘은 격자점 (m, N_1) , 즉, 표준패턴 1의 끝에서 종료되는 모든 경로에 대한 누적된 거리값들을 추적해야만 한다. 그런 경로에 해당하는 m값의 범위는

$$m_1(1) < m < m_2(1)$$

로서 이범위는 직선 $n=N_1$ 과 워핑함수의 상한 및 하한 제약직선과의 교점들에 의해 결정된다.

$$m_1(L) = \min_{1 \leq v \leq V} [m_{v1}(L)] \tag{3-2}$$

$$m_2(L) = \min_{1 \leq v \leq V} [m_{v2}(L)] \tag{3-3}$$

3.2 Level Building 알고리즘

레벨 L의 종료영역 $m_1(L) < m < m_2(L)$ 내의 각 m 값에 대해서 다음 세가지지를 만족해야 한다.[3]

(1) 누적된 최소 거리값 $\tilde{D}_1^B(m) = \min_{1 \leq v \leq V} [\tilde{D}_1^V(m)]$

여기서 $\tilde{D}_1^V(m)$ 은 V번째 표준패턴에 대해 레벨 1에서 입력패턴의 n번째 프레임에서 종료되는 누적 거리값을 의미한다.

(2) 최적 표준패턴(reference), $W_1(m) = \operatorname{argmin}_x \min_{1 \leq v \leq V} [\tilde{D}_1^V(m)]$

여기서 $\operatorname{argmin}_x [f(x)]$ 는 f(x)를 최소화 시키는 x 값을 말한다.

(3) 귀환점(backtracking pointer) $\tilde{F}_1^B = \tilde{F}_1^{W_1(m)}$

여기서 $\tilde{F}_1^V(m)$ 은 m번째 프레임에서 끝래임에서 끝나는 경로들 중 최적경로가 시작되는 입력패턴의 프레임이 어느 것인가를 지시한다. 이 포인트는 기본적으로 그 이전 레벨에 각 경로가 끝나는 프레임을 추적한다. 레벨 1=1에 대해서는 모든 m에 대해서 $\tilde{F}_1^B(m) = 0$ 이되며 이는 모든 경로가 입력패턴의 첫 프레임에서 시작되기 때문이다.

IV. 제안된 One Stage MSVQ/DP 인식 알고리즘

4.1 개요

본 연구에서는 연속음 인식에 있어서 중요한 의미를 가지는 기준패턴의 생성을 MSVQ로 작성함으로써 전자계산기의 기억용량을 효율적으로 사용할 수 있게 하였고 기존의 연결단어 인식 방법인 One Stage DP를 연속음 인식에 도입하여 인식단계의 수행시간

을 상당히 단축 시킬 수 있는 One Stage MSVQ/DP를 제안하여 인식실험을 수행 하였다. One Stage DP 알고리즘은 time warping 경로를 하나의 지표(index)로써 파라메타화(parameterizing)함으로써, time warping 경로를 직접 최적화 범위의 한 함수로 취급하는데 기초한다.[8][9]

4.2 알고리즘의 기초단계

- (1) $i=1, \dots, N$ 개의 시간 프레임(time frame)으로 구성된 마지의 단어열의 입력
- (2) 입력 단어열에 대응되는 기준패턴은 단음절로 구성된 K개의 표준템플릿(template)의 집합으로 구성된다. 즉, 하나의 템플릿마다 MSVQ에 의해 각 section 별로 구한 codeword를 생성한다.
- (3) 템플릿 k의 시간 프레임은 $j=1, \dots, J(k)$ 로서 표시되고 이때 $J(k)$ 는 길이로 모든 템플릿의 길이가 같다.
- (4) 인식목표는 입력단어열과 가장 잘 정합(matching)되는 템플릿들의 연결, 즉 $q(1), \dots, q(R)$ 을 결정하는 것이다. 여기서 $q(1), \dots, q(R)$ 을 완전히 정합된 표준패턴 단어열(super reference pattern)이라고 한다.

격자점(i, j, k)은 시험패턴의 시간 프레임 i와 템플릿 k의 시간 프레임 j로 정의되며 소구간 거리계산(i, j, k)과 관련된다. 수식적으로 경로 W는 격자점 $C(1), 1=1, 2, \dots, L$ 의 열로서 주어진다.

$$C = (C(1), C(2), \dots, C(L)) \tag{4-1}$$

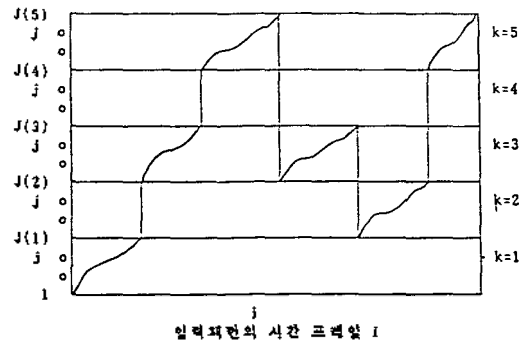


그림 4.1 One Stage MSVQ/DP 알고리즘에서 연속음의 입력과 일련의 표준 템플릿들 사이의 최적경로
Fig. 4.1 Warpaint path regions in a proposed One Stage MSVQ/DP

본 연구에서 제안한 One Stage MSVQ/DP방법의 기본개념이 그림 4.1에 나타나 있다.

여기서 정합과정은 주어진 경로에 따른 소구간 거리들의 합으로 삼는다. 즉,

$$\min_c \sum_{l=1}^i d(C(l)) \quad (4-2)$$

와 같이 모든 허용된 경로에 대한 전체 거리의 최소화 문제라고 말할 수 있다. 두 형태의 천이법칙에 대한 것이 그림 4.2에 나타나 있다.

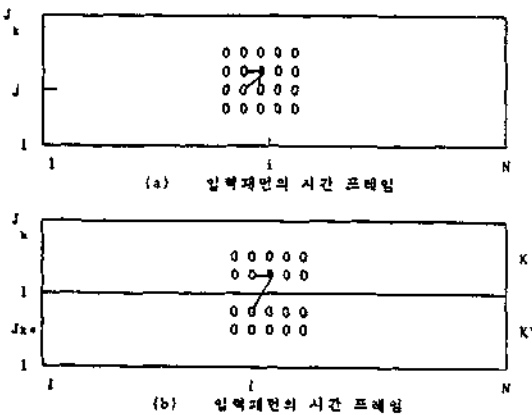


그림 4.2 One Stage MSVQ/DP 알고리즘에서의 두가지 천이법칙의 예

- (a) 템플릿 내부의 천이법칙
- (b) 템플릿 경계의 천이법칙

Fig 4.2 Transition rules in the One Stage MSVQ/DP
(a) Within-template transition rules
(b) Example of between template transition rules

만일, 그림에서 $C(i) = (i, j, k), j > 1$ 이면

$$C(i-1) \in \{(i-1, j, k), (i-1, j-1, k), (i, j-1, k)\}$$

또는, $C(i) = (i, j, k), j = 1$ 이면

$$C(i-1) \in \{(i-1, 1, k), (i-1, J(k^*), k^*): k^* = 1, \dots, K\} \quad (4-3)$$

최적의 원리를 일반화 하기위해 격자점 (i, j, k) 에서 임의의 경로를 따르는 최소의 누적거리 $D(i, j, k)$ 는 다음과 같다.

(1) $j > 1$ 인 경우에 템플릿 내부의 식

$$D(i, j, k) = d(i, j, k) + \min\{D(i-1, j, k), D(i-1, j-1, k), D(i, j-1, k)\} \quad (4-4)$$

(2) $j = 1$ 인 템플릿 경계의 식

$$D(i, 1, k) = d(i, 1, k) + \min\{D(i-1, 1, k), D(i-1, J(k^*), k^*): k^* = 1, \dots, K\} \quad (4-5)$$

이 알고리즘의 최종목표는 미지의 단어열을 결정하는 것이다. 따라서 귀환점(back pointer) $B(i, j, k)$ 는 격자점 (i, j, k) 에 이르는 경로로부터 이전단어의 마지막 프레임의 값을 기억해야 한다.

4.3 One Stage MSVQ/DP 알고리즘

단계 1. 초기화 $D(1, j, k) = \sum_{n=1}^j d(1, n, k)$

- 단계 2. a) $i = 2, \dots, N$ 에 대해서 $2b-2e$ 까지 단계를 수행
- b) $k = 1, \dots, K$ 에 대해서 $2c-2e$ 까지 단계를 수행
- c) $D(i, 1, k) = d(i, 1, k) + \min\{D(i-1, 1, k), D(i-1, J(k^*), k^*): k^* = 1, \dots, K\}$
- d) $j = 2, \dots, J(k)$ 에 대해서 $2e$ 단계를 수행
- e) $D(i, j, k) = d(i, j, k) + \min\{D(i-1, j, k), D(i-1, j-1, k), D(i, j-1, k)\}$

단계 3. 누적된 거리에 대한 배열 $D(i, j, k)$ 를 사용하여 최소거리값을 가지는 템플릿의 끝 프레임에서 격자점으로 최적의 경로를 추적(trace back)한다.

이 알고리즘의 최종목표는 미지의 단어열을 결정하는 것이다. 따라서 단계3에서 귀환점(back pointer) $B(i, j, k)$ 는 격자점 (i, j, k) 에 이르는 경로로부터 이전단어의 마지막 프레임의 값을 기억해야 한다. 배열 $T(i)$ 는 시험패턴의 각 프레임 j 에 대하여 그 마지막 프레임 $J(k)$ 에서 최소의 누적된 거리값을 가지는 템플릿의 지표(index)를 기억하는데 사용한다.

$$T(i) = \underset{1 \leq k \leq K}{\operatorname{argmin}} \{D(i, J(k), k) = D(J(k), k): k = 1, \dots, K\} \quad (4-6)$$

또한 배열 $F(i)$ 는 표준패턴의 단어 경계가 되는 위치에서 시험패턴의 프레임 i 를 기억하게 된다.

$$F(i) = B(i, J(T(i)), T(i)) \quad (4-7)$$

이 방법은 한 단어당 인식시간의 계산량이 입력단어의 갯수에 상관없이 실행될 수 있다는 점에서 패턴정

합에 대한 문제를 여러 레벨(level)들로 나누어 인식하는 다른 방법들 보다 훨씬 간편하고 효율적이다.[8]

V. 인식실험

본 연구는 인식 실험 대상어휘로 대학교의 일반적인 행정 부서명 40개를 선정하였다. 음성 데이터는 방음장치가 되어 있지 않고 컴퓨터 잡음이 있는 보통의 환경에서 다이내믹 마이크를 통해 저장되었다. 또한, 화자로는 3명의 남성(20, 30, 40대 남성)을 선택하였으며, 각 화자마다 대학교 행정 부서명 40개를 3번씩 발음하였고, MSVQ에 의한 codeword, 즉, 기준패턴을 생성하는데는 각각 2번씩 발음한 것을 사용하였으며, 나머지 각각 1번씩 발음한 것으로 인식 실험을 수행하였다. 여기서 사용된 대학교 행정부서명의 명칭은 표 5-1과 같다.

5.1 인식 시스템 구성

본 연구에서 실험한 음성 인식 시스템은 그림 5.1과 같으며, 실험에 사용된 대학교 행정 부서명 데이터는 마이크에 의해 직접 입력된 음성 신호로써, 샘플링 주파수를 8 kHz로 하였으며, 3.5 kHz 저역 여파기를 통과한 후 12 Bit A/D 변환기를 거쳐 음성 데이터를 구하였다. 그리고, 시작점과 끝점을 검출한 후, LPC 계수보다 LPC 칩스트림 계수를 특징 벡터로 사용하는 것이 더 좋다는 고찰에 따라 특징 파라미터로는 LPC 칩스트림을 사용하였으며, 이것을 이용하여 MSVQ[5][7]에 의해 각 단어마다 codebook을 가진 기준패턴을 생성하여 저장하고 본 연구에서 제안한 One Stage MSVQ/DP에 의해 인식실험을 수행 하였다.

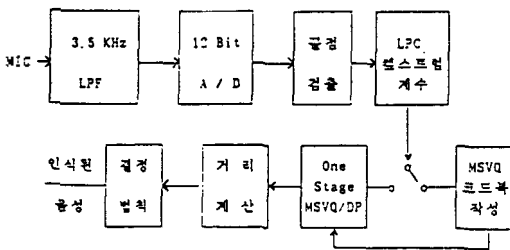


Fig 5.1 인식 시스템의 구성도

그림 5.1 Block diagram of recognition system

표 5.1 연속어의 단음절 기준패턴 목록

Table 5.1 Reference Pattern List of syllables

1	총(chong)	16	도(do)	31	이(ii)	46	시(si)
2	무(mu)	17	서(seo)	32	공(kong)	47	설(seal)
3	과(kwa)	18	박(bak)	33	정(jung)	48	열(yeal)
4	기(ki)	19	물(mul)	34	치(chi)	49	람(ram)
5	혁(heik)	20	신(shin)	35	법(buk)	50	회(hye)
6	교(kyo)	21	문(mun)	36	상(sang)	51	환(hwan)
7	학(hak)	22	사(sa)	37	축(chuk)	52	당(dang)
8	생(seng)	23	방(bang)	38	농(nong)	53	직(jik)
9	관(kwn)	24	송(song)	39	번(bum)	54	비(bi)
10	리(ri)	25	국(kuk)	40	화(hwa)	55	우(wu)
11	적(juk)	26	전(jun)	41	의(eui)	56	채(chae)
12	장(hong)	27	산(san)	42	실(sil)	57	군(kun)
13	홍(hong)	28	소(so)	43	습(seup)	58	단(dan)
14	보(bo)	29	대(dae)	44	수(su)		
15	경(kyung)	30	원(won)	45	위(wi)		

표 5.2 실험 대상의 연속음

Table 5.2 Continuous Speech of the experiment object

1	총무과	11	도서관	21	법과대	31	시설과
1	기획과	12	박물관	22	상경대	32	수서과
3	교학과	13	신문사	23	축산대	33	열람과
4	교무과	14	방송국	24	농과대	34	총장실
5	학생과	15	전산소	25	사범대	35	학생회
6	관리과	16	대학원	26	문화대	36	교환실
7	학적과	17	문과대	27	의과대	37	당직실
8	장학과	18	이과대	28	실습실	38	경비실
9	홍보과	19	공과대	29	전기실	39	우체국
10	경리과	20	정치대	30	수위실	40	학군단

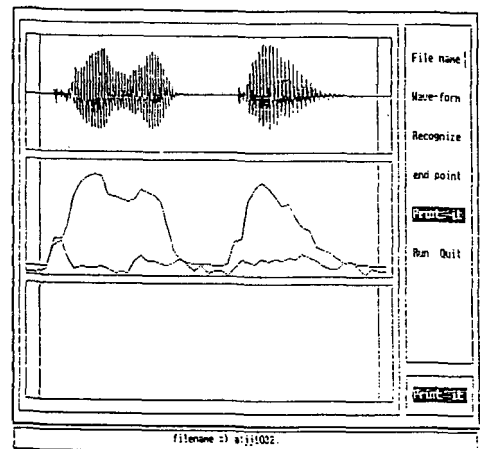


Fig 5.2 /교학과에 대한 끝점검출

그림 5.2 Endpoint detection of/kyo.hak.kwa/

표 5.3 인식실험 결과표

Table 5.3 Result table by the Recognition Experiment

화자 방법	화자1	화자2	화자3	총인식율
LBDTW	82.5%	72.5%	83.3%	
One Stage	87.5%	89.9%	85%	87.5%
One Stage MSVQ/DP	92.5%	95%	87.5%	91.6%

5.2 인식실험 및 고찰

본 연구의 인식 대상 어휘는 실시간 음성 다이얼링 인식 시스템용 구현 하기위한 것으로 대학교 행정 부서명 40개를 선정 하였다. 본래 이들 LBDTW와 One Stage DP 알고리즘은 연결음 인식 알고리즘으로 사용된 것인데 특별하게 연속음을 인식하는 알고리즘이 대두되지 않은 상태여서 이 알고리즘은 지금까지 연결음 또는 연속음 인식 알고리즘으로 사용 되어 왔다. 그러나 DTW 방법으로 인식율은 향상 되어 왔지만 인식시간이 길어지는 단점이 있었다. 그러한 이유로 본 실험에서는 첫번째로 데이터를 압축시키는 방법으로 VQ방법을 이용하였고 더우기 연속음의 시간적 변화율을 줄이기 위해 한 단어당 6개의 section으로 나누는 MSVQ 방법을 이용해 연속음의 단음절로 기준패턴을 생성하였으며 모든 기준패턴들이 레벨마다 매칭되고 누적된 거리값이 최소가 되는 단어열을 인식하는 LBDTW 방법보다 한 영역상에 모든 기준패턴을 올려놓고 동시에 매칭 시켜 나가는 효율성 높은 One Stage DP방법을 도입해 연속음을 대상으로 One Stage MSVQ/DP 방법을 제안해 인식실험을 수행하였다.

본 연구에서 기준패턴을 작성하는 방법에 따라 인식율과 인식시간에 미치는 영향이 대단히 큼을 알 수 있었다. 기존의 인식방법인 LBDTW나 One Stage의 기준패턴을 음성의 프레임으로 프레임마다 길이의 차이를 가지고 있지만 VQ로 작성된 기준 패턴은 음절당 6개의 section으로 같아서 인식을 위한 계산이 간편하고 빠르다. 하지만 인식율을 높이기 위해서는 단어당 많은 토른을 놓고 통계적인 기준패턴을 설정하는 것이 바람직 하다고 생각된다.

VI. 결 론

본 연구는 기준패턴의 기억용량을 작게 하기 위해서 VQ방법을 이용했고 더욱 효율적인 방법으로 시간적인 변화를 고려해 6개의 section으로 구분해 code-book을 구성했으며 one stage MSVQ/DP를 이용해 인식실험을 수행하였다. 음절당 6개의 section으로 설정한 것은 section수가 너무 적으면 음성정보가 손실되고 너무크면 계산이 많아지므로 인식속도가 느려지기 때문이다. 전체적인 인식실험은 비교를 위해 같은 데이터 같은 조건상에서 기존의 인식방법인 LBDTW 인식 시스템, stage DP 인식 시스템과 함께 인식실험을 하였다. 실험결과에서 알 수 있듯이 인식 시간과 기억용량을 고려한 MSVQ에 의한 기준패턴 작성법이 다른 우수함을 알 수 있다. 본 연구에서 제안한 One Stage MSVQ/DP에 의한 총인식율은 91.6%이다.

참 고 문 헌

1. T. K. Vintysyuk, "Element-Wise recognition of continuous speech composed of words from a specified dictionary," Kiber netika, Vol.7, pp. 133-143, 1971.
2. C. S. Myers and L. 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.
3. 정의봉, 김순협, 이종악, "L.B 알고리즘을 이용한 한국어 숫자음 인식," Vol. 4, No.5, pp. 15-19, 한국음향학회, 1988.10.
4. Kuk-Chin Pan, Frank K. Soong and L. R. Rabiner, "A Vector Quantization-Based Preprocessor for Speaker-Independent Isolated Word Recognition," IEEE Trans. of Acoustics, Speech, Signal Processing, Vol. ASSP-33, No. 3, Aune, 1985.
5. D. K. Burton and J. E. Share, "Speaker-Dependent Isolated Word Recognition using Speaker-Independent Vector Quantization Codebooks Augmented with Speaker-Specific Data," IEEE Trans. of Acoustics, Speech, and Signal Processing, ASSP-33, No. 2, pp 440-443, Apr. 1985.
6. Shikano, K. and Kohda, M., "On the LPC Distance Measures for Vowel Recognition in Continuous Ut-

- terance," Institute of Electrical and Communication Engineers of Japen, Trans. on D. J, 63-D, May, 1980.
7. 안태욱, 김순협, "개선된 MSVQ 인식 시스템을 이용한 한국어 음성 인식," 한국통신학회, 제27권 제12호, PP. 1-10, 1991.3.
 8. Herman Ney, "The use of One-Stage Dynamic Programming Algorithm for connected word recognition," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. Assp-32, No. 2, APR.1984.
 9. 정의봉, "One Stage MSVQ/DP를 이용한 연속음 인식에 관한 연구," 박사학위논문, 건국대학교 대학원, 1992.2.

▲정의봉



제10권 2호 참조
전북산업대학교 정보통신공학과
전임강사