

# LPC Cepstrum 과 집단화를 이용한 한국어 고립단어 인식에 관한 연구

## A study on Korean isolated word recognition using LPC cepstrum and clustering

\*김 진 영 (Kim, J. Y)

\*\*성 광 모 (Seong, G. M.)

### 요 약

본 논문은 화자독립 고립단어 인식에 있어서 LP모델의 문제점과 그 해결 방안으로서 cepstrum 영역에 있어서 lifter를 이용한 해결에 대해서 고찰하였다. 한편, 각 인식 단어의 기준 패턴을 구하기 위한 방법으로서 집단화의 방법에 대해 논하였다. 집단화의 방법으로는 UWA방법과 K-iteration 방법을 변형시킨 KMA 방법을 제시 비교하였다. 인식실험결과 정현과 lifter와 KMA의 집단화 방법을 사용하였을 때 95%의 최고 인식률을 보였다.

### ABSTRACT

In this paper, the problem of LP-model and it's solution by liftering in cepstrum domain are investigated in speaker independent isolated-word recognition. And, clustering technique is discussed for obtaining the reference template. KMA (K-means iteration with average) method, which is transformed from UWA method and K-iteration method, has been suggested and compared with each other for clustering, the result of recognition experiments shows max. 95% recognition rate when raised-sign lifter and KMA clustering method is applied.

\* 서울대 전자공학과 석사과정

\*\* 서울대 전자공학과 교수

## I. 서론

인간의 음성신호 처리기술은 디지털 컴퓨터의 발달과 통신의 고급화에 힘 입어 최근 급격히 발달하였다. 이에 따라 인간의 음성에 대한 인식 즉, 사람과 기계와의 대화가 가능하게 되었다. 1960년 후반부터 음성인식에 대해 많은 연구가 수행되어 왔는데 특히, 고립단어 인식분야에 있어서 많은 성과가 있어 왔다. 고립단어인식의 방법으로서 여러가지가 제안되어 왔는데 그중 패턴매칭(pattern matching)의 방법이 가장 인식률이 높은 것으로 알려져 있다. 그런데 패턴매칭 방법에 있어서 가장 중요한 것은 인식 파라미터의 설정과 기준 패턴의 설정 방법에 있겠다.

지금까지 알려진 인식 파라미터는 크게 LPC와 filterbank의 계수로 나눌 수 있고 LPC에 의한 인식이 더 나은 인식률을 보여주고 있다. 그러나, LPC에 의한 스펙트럼 추정이 화자간의 고유의 성질을 완전히 제거하지 못하여 화자독립 인식기의 개발에 있어 장애요소가 되어 왔다<sup>(1)</sup>. 따라서 LP 계수에 대해 변형이 요구되는데 LP 계수를 각각 독립적으로 다룰 수가 없으므로 각 계수가 독립적이 되

도록 변화시켜야 한다. 이를 위해 각 계수가 독립적인 cepstrum영역으로의 변환을 하여 조작을 가하게 된다. 한편, 기준패턴을 구하기 위해 집단화의 기법이 요구되는데, Rabiner 등에 의해 여러가지가 제안되어 왔다<sup>(3,4)</sup> 본 논문에서는 Rabiner가 제안한 UWA 방법과 K-iteration 방법에 평균화 기법을 도입한 KMA 방법을 제시하고 비교하고자 한다.

## II. 음성 인식 시스템의 모델

현재까지의 음성인식 시스템의 모델은 여러가지가 있으나 본 논문에서는 기본적인 인식 시스템으로서 인식률이 가장 높은 것으로 알려진 패턴 정합의 방법을 이용하기로 하였다. 따라서 제안한 고립단어 인식 시스템은 그림 2-1과 같다.

마이크를 통하여 입력된 신호는 4 KHz 저역여파기를 통과한 후 8 KHz의 16bit A/D 변환을 하여 음성 구간을 추출하고 분석 프레임이 유성을 또는 무성음인지를 분류하게 된다. 이렇게 하여 음성 부분만을 찾아낸 뒤 특성 파라미터를 얻기 위하여 스펙트럼 분석을 선형예측(linear predictor)의 방법으로 예측 계수를 구한 후, 이것을 LPC cepstrum 계수로 변환하게 된다. 이것은 특성 파라미터의 변형

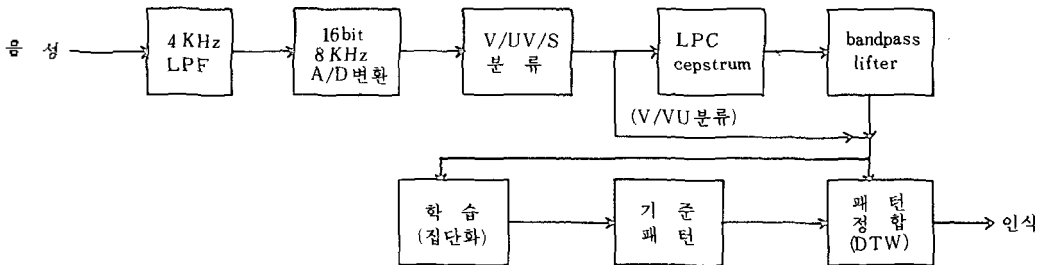


그림 2-1 제안한 고립단어 인식시스템.

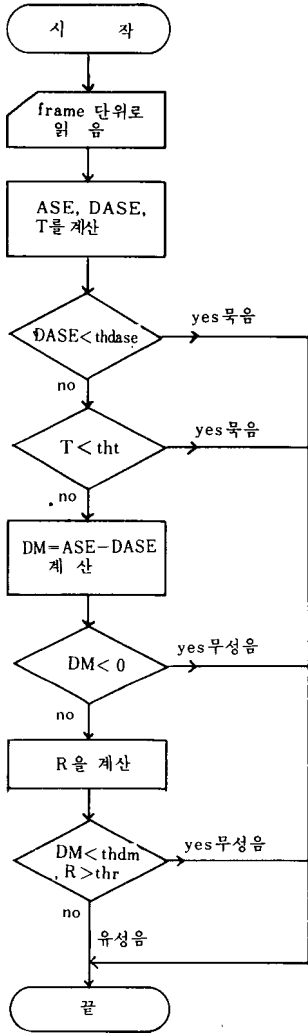


그림 2 - 2 유성음 / 무성음 / 목음 분류알고리즘.

을 용이하게 하기 위한 것이다. 얻어진 cepstrum 계수는 lifter를 통과함으로써 최종적인 특징 파라미터가 된다. 한편 단어의 기준 패턴을 만들기 위하여 얻어진 데이터들로 학습과정을 행하는데 이것은 집단화의 방법을 이용한다. 이렇게 하여 만들어진 기준 패턴 시험패턴의 유사도를 측정하기 위해 DTW라는 동적 프로그램(dynamic program)의 기법을 사용하여 각 단어의 기준 패턴과 유사도를 측정,

가장 유사도가 큰 단어로 인식하게 된다. 한편, 많은 단어를 인식하는데 있어서 그 단어들의 특징에 따라 비교하는 단어의 수를 줄일 수 있는데 본 논문에서는 무성음의 유무 또는 무성음이 있다면 그 위치를 사용하였다. 예를들어 공, 일, ...구까지 인식 한다면 무성음이 존재하는 단어가 반이므로 계산량을 반으로 줄일 수 있다. 여기서 중요한 것은 정확히 무성음의 구간을 찾는 것이 아니라 그 유무 성만을 알면 된다는 것이다. 그 방법으로서 level crossing와 DPCM을 사용한 유성음 / 무성음 / 목음 분류 알고리즘을 사용하였다<sup>(5)</sup>. 그림 2 - 2는 그 분류 알고리즘을 보여주고 있다. 여기서 사용 파라미터는

ASE = (프레임의 절대값 에너지)

BASE = (DPCM 후의 절대값 에너지)

T = (목음 구간에서 정해진 level을 넘는 peak의 합)

이다. 위의 파라미터와 통계적인 경계값으로 유성음 / 무성음 / 목음 분류를 한다.

### III. LPC cepstrum

cepstrum 분석은 본래 지진과 연구에 이용되었던 방법으로 그림 3 - 1과 같은 나타낼 수 있다. 입력된 음성 신호의 전력밀도 스펙트럼을 P(w)라고 할 때 다음과 같이 표시될 수 있다.

$$P(w) = |G(w)|^2 |H(w)|^2 = |V(w)|^2 |E(w)|^2 \quad (1)$$

단, S(w) : 신호의 스펙트럼

G(w) : 음성의 발생 신호

V(w) : 이상적인 음성의 발생 신호 (impulse train)

$H(w), E(w)$  : 스펙트럼의 envelope

과 같다.

여기서 quefrency가 제한된 cepstrum  $C(t)$ 는 다음과 같이 정의된다.

$$C(t) = l(t) \cdot C^{\circ}(t) \quad (2)$$

$$= \begin{cases} \int_0^{\infty} e^{j\omega t} \log P(\omega) d\omega & \text{if } (t \leq T); \\ 0 & \text{if } (t > T). \end{cases}$$

$$l(t) = \begin{cases} 1 & \text{if } (t \leq T); \\ 0 & \text{if } (t > T). \end{cases}$$

여기서  $C(t)$ 를 이용하여 분석하는 것을 cepstrum 분석이라고 한다.

### 1. LPC cepstrum

위에서 cepstrum을 구하기 위하여서 입력 신호의 전력밀도 스펙트럼을 이용하였다. 그런데 스펙트럼 추정 이론에 의하면 스펙트럼은 선형예측에 의하여 얻을 수가 있다. 음성의 coding 방법에 있어서 그러한 방법을 LPC(linear predictive coding)라고 한다. LPC에 의하여 구하여진 스펙트럼을  $s\sqrt{\alpha}/A(z)$ 이라고 하면

$$S(\omega) = \sqrt{\alpha}/A(z) \quad (3)$$

$$\text{단, } A(z) = \sum_{i=0}^M a_i z^{-i}$$

$M$  : LPC의 차수

따라서  $P(\omega)$  대신에 위의 스펙트럼 근사를 이용하여 cepstrum을 구하여 보면

$$C_n = \frac{1}{2\pi} \int_0^{2\pi} \log |S(\omega)| e^{jn\omega} d\omega. \quad (4)$$

이 된다. 이것을 푸는 방법을 간단하게 반복적인 형태로 구할 수 있도록 주어져 있고 그 식은 다음

$$-nc(n) - na_n = \sum_{k=1}^{n-1} (n-k) c^{\circ}(n-k) a_k, \quad 1 \leq n \leq p. \quad (5)$$

위와같이 cepstrum을 LPC의 스펙트럼을 이용하여 구하는 것을 LPC cepstrum이라고 한다.

### 2. LPC(LPC cepstrum)의 문제점과 해결

음성 신호의 LP 모델은 pitch와 같은 음성 발생 요소에 어느 정도 둔감하기는 하나 전혀 상관성이 없는 것은 아니다. 또한 all-pole 모델이라는 제약조건 때문에 음성인식에 바람직하지 못한 스펙트럼성분을 만들어 낸다<sup>(1)</sup>. 예를 들어 zero가 없으므로 스펙트럼 영역에서 각 첨두치들의 폭(band width)이 매우 작게 되어 그 중심 주파수가 조금만 바뀌어도 매우 큰 차이를 가져오게 된다. 즉 각 개인의 pitch에 따라 영향을 받게 된다. 따라서 변형을 가해야 하는데 LPC의 계수는 조작에 어려움이 있다.

왜냐하면 각 계수가 독립적이지 못하기 때문이다. 따라서 앞에서 소개한 cepstrum 변환을 하게되는 것이다. quefrency 영역에 보면 저주파 성분의 pitch가 높은 quefrency에 영향을 주게 된다. 한편 음성이 발생하는 과정에서 음파가 성도를 따라서 진행하게 되는데 이때 고주파 성분이 저주파 성분보다 많은 변화를 겪게 된다. 그런데 각 개인의 성도는 사람마다 다르므로 고주파 성분의 변화는 저주파보다 개인에 따른 변화가 크다. 이러한 영향은 낮은 quefrency 성분에서 나타난다. 또한 LP스펙트럼은 glottal의 모형과 glottal의 duty cycle의 성분을 포함하고 있는데 이것은 낮은 quefrency에 영향을 준다. 한편 LP분석에 있어서 바람직하지 못한 스펙트럼의 변화는 신호 스펙트럼에 스펙트럴 notch나 zero가 나타나는 경우이다. 즉 all-pole의 LPC분석은 필터링후 스펙트럼의 평탄성이 최대가 되도록 함으로 낮은 quefrency에서 큰 변화를 일으킨다. 따

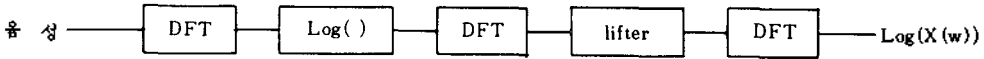


그림 3 - 1 음성의 cepstrum 분석.

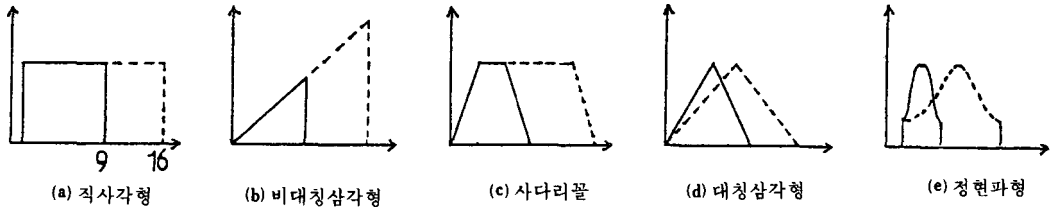


그림 3 - 2 lifter 의 예

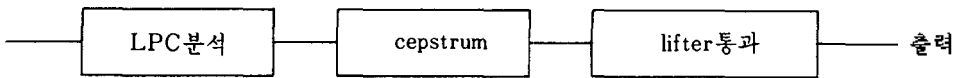


그림 3 - 3 lifter 를 사용한 분석.

라서 cepstrum의 영역에서 lifter를 사용해야 할 필요성이 있다. lifter로는 여러가지가 있겠으나 그림 3 - 2와 같은 것들이 제안되어 있다<sup>(1)</sup>. 그림3 - 2와 같은 lifter를 사용하여 음성인식을 하는 시스템은 그림 3 - 3과 같이 나타낼 수 있다.

그림.3 - 4는 '구'의 'T'에 해당하는 부분의 스펙트럼, 8차의 LPC모델 스펙트럼, 9, 12, 15차의 cepstrum 스펙트럼, 그리고 그것의 lifter를 거친 후의 스펙트럼이다. 그림에서 보듯이 첨두치가 완곡해진 것을 볼 수 있다. 또한 LPC cepstrum의 차수가 증가할수록 골(local minima)와 산(local maxima)이 강조되어지는 것을 알 수 있다. Lifter를 통과한 후의 스펙트럼은 그 스펙트럼의 완만성은 유지하면서 formant 부분이 강조된 것을 볼 수 있다. 여기서 사용한 lifter는 정현파형(raised sine)으로써 그 식은 다음과 같다.

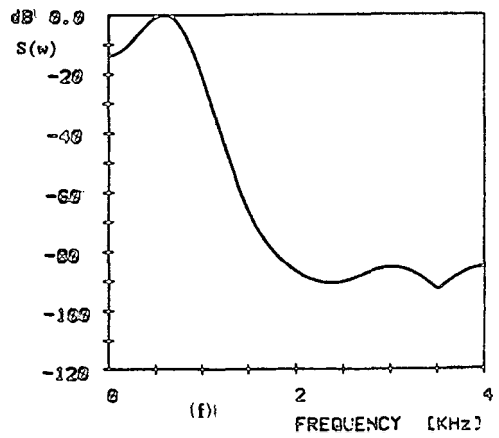
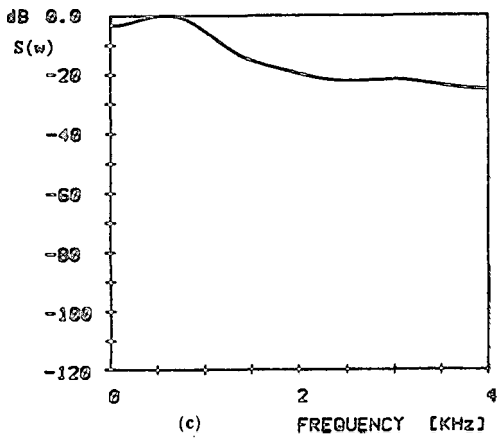
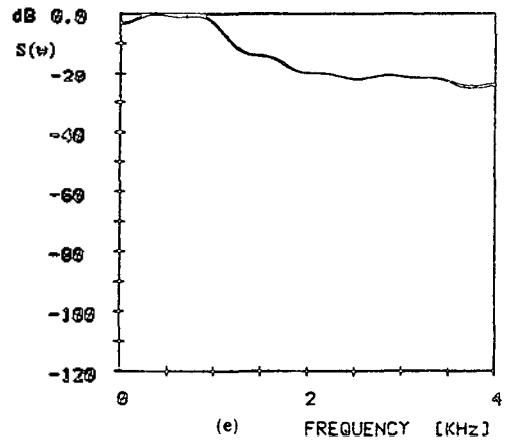
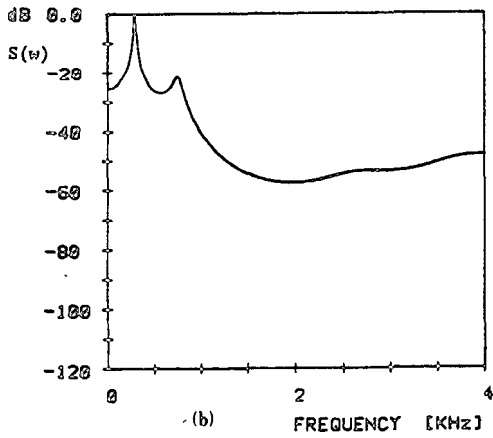
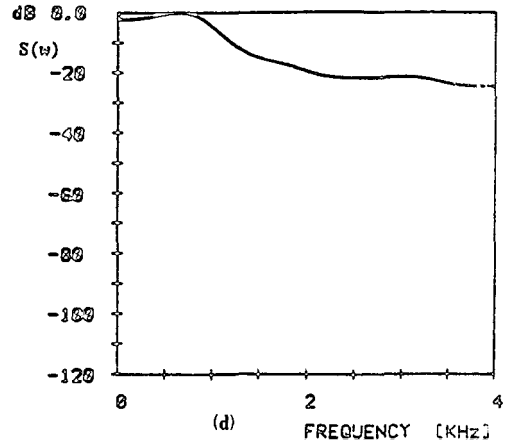
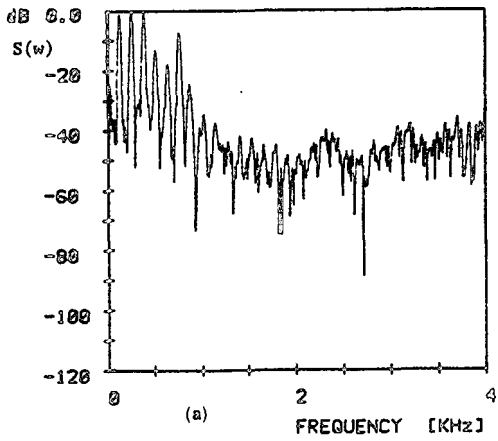
$$C(n) = 1 + h \times \sin(n \times 3.14/L) \tag{6}$$

단, L : cepstrum 차수

$$h : 0.5 \cdot L$$

#### IV. 집단화의 기법

패턴정합의 방법을 이용한 고립 단어 인식(특히, 화자독립)에 있어서 중요한 문제 중의 하나는 인식하고자 하는 단어들의 기준 패턴을 만들어 내는 것이다. 이 기준 패턴을 형성하는데 있어서 중요한 것은 한 단어에 대해 만들어진 여러 패턴 중에서 어떤 방법으로 기준 패턴(하나 또는 그 이상)을 추출할 것인가가 되겠다. 물론 여러 개의 패턴 중에서 임의의 것을 기준 패턴으로 선택하여 인식을 할 수도 있겠으나 많은 오인식을 유발하게 될 것이다. 따



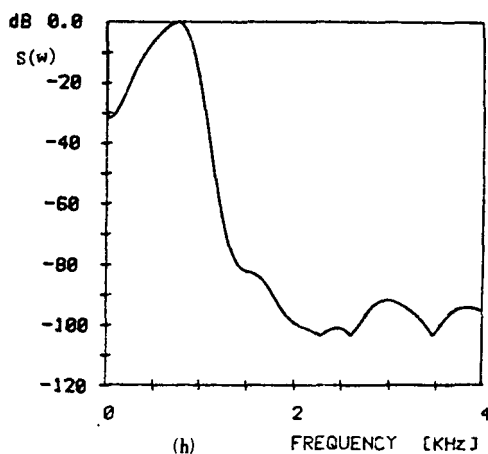
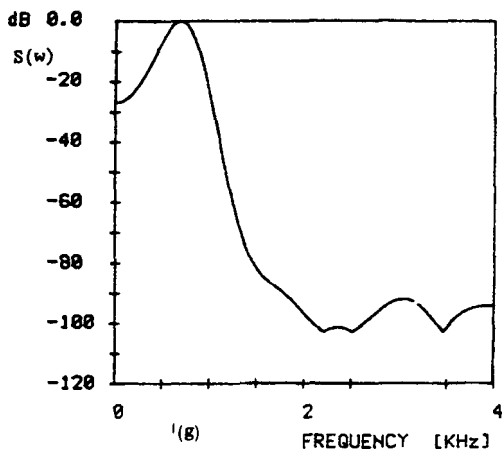


그림3-4 '우'의 스펙트럼.

(b) LPC 9차 스펙트럼 cepstrum 스펙트럼 (c) 9차 (d) 12차 (e) 15차  
lifter된 스펙트럼 (e) 9차 (f) 12차 (g) 15차

라서 집단화의 기법을 사용하여 가장 최적의 기준 패턴을 만들어야 할 것이다. 지금까지 그 기법으로서 여러가지가 제안되어 왔다. 본 논문에서는 그중 간단한 방법인 UWA와 K-means iteration의 방법을 변형시킨 KMA (K-means with average)을 제시하고자 한다.

1. UWA (Unsupervised clustering without averaging)

이 방법은 그림 3-1에 나타내었는데 집단의 결정 threshold 값만 주어지면 자연스럽게 집단의 갯수와 집단의 원소가 정하여지는 방법이다. 이 알고리즘에서 두 패턴사이의 거리 즉  $X_i$ 와  $X_j$ 의 거리는 앞 장에서 기술한 DTW방법을 이용하여 구하여진다. 한편  $d_{ij} \neq d_{ji}$ 인 경우를 고려하여 다음과 같이 거리를 정의한다.

$$d_{ij} = \frac{d_{ij} + d_{ji}}{2} = \frac{\delta(x_i, x_j) + \delta(x_j, x_i)}{2} \quad (7)$$

집단들의 집합을  $W_1, W_2, \dots, W_r$ 라고 할 때 집단화를 하고자 하는 패턴 중  $W_1, W_2, \dots, W_r$ 에 포함되지 않은 패턴들의 집합은 다음과 같이 정의된다.

$$\Omega'_{i+1} = \Omega - U'_{i+1}, w_i = \Omega'_i - w_i = \{x'_1, x'_2, \dots, x'_{q(i)}\}$$

단,  $\Omega$  : 전체 집합

$x'_i$  :  $\Omega'_{i+1}$ 의 요소

$q(j)$  : 처음  $j$ 개의 집단형성 후에 포함되지 않은 수

그러면 UWA 알고리즘은 다음과 같다.

- 1) 초기화  $j = 0$
- 2)  $\Omega'_{i+1}$  관찰 집합 중 minmax 중심(center)를 구한다. 이 minmax 중심을  $\hat{x}_{i+1}$ 라고 한다.  

$$\text{minmax} : \hat{x}_{i+1} = x'_i \ni \max_j \delta(x'_i, x'_j) \leq \min_i \max_j \delta(x'_i, x'_j)$$
- 3)  $w_i$ 들의 초기 선택 ( $k = 0$ )은 다음과 같다.

$$w_i^{(k)} = U_{i \in \Omega'_{i+1}} x'_i \ni \delta(\hat{x}_{i+1}, x'_i) \leq T$$

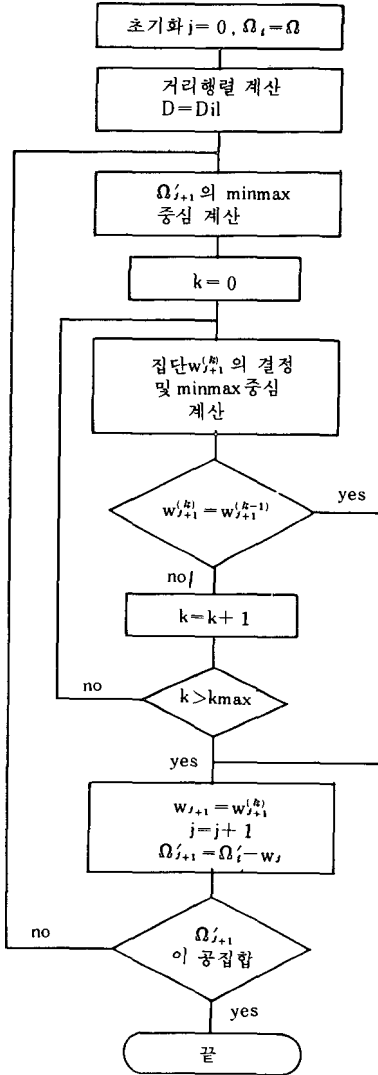


그림 3 - 1 UWA 방법

- 4)  $w_j^{(k)}$ 의 minmax 중심  $x_{j+1}$ 를 선택한다.
- 5)  $k$ 를 증가시키고 다시  $w_j^{(k)}$ 의 원소를 결정한다.  $w_j^{(k)} = w_j^{(k-1)}$  이거나  $k > KMAX$ (max iteration) 이면 완료된다. 그렇지 않으면  $j$ 를 증가시키고  $\Omega_{j+1}'$ 를 결정한다.  $\Omega_{j+1}'$ 의 집합이 공집합이 아니면 2)로

간다. 위의 check에 걸리지 않으면 stop로 가라.

위와 같은 UWA 방법은 거리 계산 처음 단계에서  $d$ 의 행렬  $D$ 를 구한 뒤 다시는 거리 계산 과정이 없으므로 빠른 계산을 통해 기준 패턴을 구할 수 있다.

2. KMA (K-means iteration with average)

K-means iteration 알고리즘은 특정 수의 집단을 찾는데 유용한 자동화된 반복 과정이다. 이 반복 과정은 분류(classification), 집단 중심의 계산, 수렴 테스트의 세가지 과정으로 이루어져 있다. 만약  $M$ 개의 집단을 찾고자 한다면 초기집단의 중심으로서  $M$ 개의 임의의 패턴을 지정해야 한다. 그러면 임의로 주어진 초기 중심들에 따라 그 성능이 저하되는 단점이 있다. 이것은 K-means 알고리즘이 local optimum 값을 찾기 때문이다. 대부분의 음성인식시스템에서는 집단의 중심을 minimax 중심으로 찾으나 여기서는 DTW에 의한 평균 기법을 도입하여, 변형된 새로운 방법 KMA를 제시하고자 한다. 그 방법은 아래와 같다.

- 1) 초기화  $X_j^{(0)} = x_i, 1 \leq i \leq M$
- 2) nearest neighbor 법칙에 의하여 패턴들을  $M$ 개의 집단에 귀속시킨다.

즉,

$$x_i \in w_k \text{ iff } \delta(x_i, x_j^{(k)}) \leq \delta(x_i, x_p^{(k)}),$$

$$1 \leq k \leq M$$

여기서,  $j$ 는 모든 패턴들의 갯수이다.

- 3)  $w_k$ 에 대하여서 DTW를 이용 평균을 취한다.
- 4), 2), 3)을 행한다. 새로운 중심과 과거 중심이 같거나 iteration 수가 주어진 값을 넘으면 완료시킨다.

위에서 평균 기법은 다음과 같은 과정을 통한다. 즉 두 패턴  $x_i$ 와  $x_j$ 를 평균하려면 DTW를 통하여 최적의 path의 warping 함수를 찾는다. 그러면,



$$a(k) = 1/2 \{x_i(i(k)) + x_j(j(k))\} \quad (8)$$

$$m=i(k), n=j(k), k=1, 2, \dots, K$$

이 된다. 이것을 기준 프레임 수로 만들기 위해서 interpolation을 하면 된다. 다른 패턴에 대해 위와같은 과정을 반복한다. 그림 3-2는 KMA 알고리즘을 보여준다.

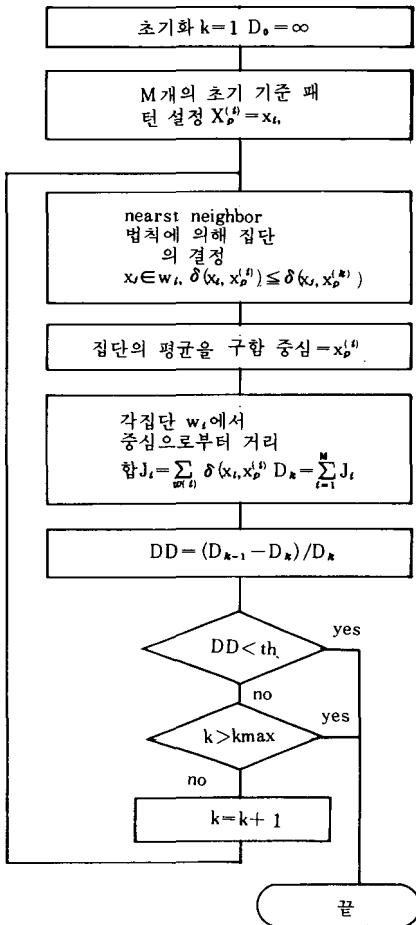


그림 3-2 KMA 방법.

## V. 실험 및 검토

### 1. 실험 데이터의 채집

한국어 고립단어 인식에 있어서 그 대상어휘는 그 쓰이는 바에 따라서 다양한 어휘선택을 할 수 있다. 그러나 본 논문에서는 인식 대상 어휘로서 음성 다이얼링(Voice-dialing)을 위한 한국어 숫자음을 사용하였다. 데이터의 수집은 4명의 화자가 각각 3번씩 발음하여 총 120개의 숫자음을 얻도록 하였다. 마이크를 통하여 입력된 음성은 4 KHz의 저역여파기를 통과한 후 16bit 8 KHz A/D변환을 하였다. A/D 변환을 한 신호는 200샘플단위의 프레임으로, 각 프레임은 100샘플씩 중첩하여 나누었다. 파라미터는 8차의 LPC 분석에 12차의 cepstrum 변환을 하여 추출하였다. 또한 각 단어당 프레임 수는 interpolation을 이용하여 25프레임으로 하였다. 이것은 숫자음의 평균 길이에 해당한다.

### 2. UWA와 KMA의 비교

UWA와 KMA의 두 집단화 방법을 비교하기 위해 cepstrum 계수를 정현파의 lifter를 사용한 파라미터로 추출하여 실험하였다.

여기서는 단순히 두 집단화 알고리즘을 비교하고자 하였으므로 학습패턴들과 시험 패턴들을 같게 하였다. 표 5-1은 기준 패턴이 하나인 경우와 둘인 경우의 인식률을 나타낸 것이다. 표 5-1에 의하면 KMA 방법이 UWA 방법에 비하여서 11.5% 정도 인식률이 나은 것을 알 수 있다. 인식 실험시 원인을 살펴본 결과 UWA 방법에 있어서의 오인식은 화자에 대한 의존성 때문에 생긴 것으로 보인다. 즉, UWA의 방법은 minmax에 의하여 기준 패턴을 구하는데 이것이 어느 특정화자의 발음한 패턴이기 때문이다. 따라서 “일”이란 음성을 인식하고자 할 때 “일”의 기준 패턴이 다른 화자의 것이고, “이”

의 기준 패턴이 같은 화자의 것이라면 “이”로 인식하는 경향을 보여 주었다. 그러나, KMA 방법은 집단화 알고리즘상에서 평균을 취하는 과정이 있어 화자에 의한 의존성을 UWA 방법에 비하여 잘 제거해주는 것임을 알 수 있었다. 그러므로 KMA 방법이 UWA 방법에 비하여 인식이 높게 나타난 것이다. 만약, UWA 방법도 평균을 취하는 과정이 삽입된다면 인식이 향상될 것으로 생각된다. 한편, 일반적으로 알 수 있듯이 각 단어당 기준패턴의 수가 증가함에 따라 인식이 올라가는 것을 볼 수 있다. 단, KMA 방법에 있어서는 UWA 방법에 비하여서 크게 향상되지는 않았는데, 이것은 평균화 과정에 의해 화자 의존성이 제거되어 어느 정도 포화 상태에 이르렀다고 볼 수 있다.

위에 논의될 인식 실험에서는 UWA 방법보다 우수하다고 판정된 KMA 방법을 사용하고 기준 패턴의 수는 1개로 고정하여 실험을 하겠다.

표 5-1. UWA와 KMA의 비교(DTW window = 6).

	UWA	KMA
1개	81.2%	93.9%
2개	89.7%	94.9%

## 2. Liftering에 의한 인식 비교

2장에서 LP모델의 문제점과 cepstrum 영역에서 lifter를 사용한 그 해결에 대해 고찰해 보았다. 그림 2-2에서 lifter의 여러가지 형태를 보인 바 있다. 그림에서 삼각형과 정현파의 lifter는 서로 유사한 것으로 간주할 수 있다. 따라서 직사각형-실제 lifter를 통과하지 않는 것과 동일-과 정현파, 비대칭 삼각형 모양의 lifter에 대하여 인식 실험을 하였다. 여기서 비대칭 삼각형에 대해 인식 실험한 것은 이 모양의 lifter가 특별한 물리적 의미가

있기 때문이다. 이 모양의 lifter를 통과한 후 Uclidean 거리를 측정하는 것은 Log 스펙트럼을 미분을 취한 뒤 거리 제산을 하는 것과 같다. 이러한 것은 백색 잡음이 있거나 formant 부분이 잡음보다 에너지를 경우 잡음의 영향을 거리 계산에서 제거하는 효과가 있다. 이를 Spectral-slope 거리라 한다.

인식실험의 결과를 표 5-2과 표 5-3에 보였다. 우리가 인식 실험을 할 때 학습의 화자와 시험 화자가 같을 때 closed-test, 다를 때 open-test라고 한다. 여기서 기준 패턴은 하나이다.

표 5-2에 의하면 직사각형의 lifter의 경우에 비하여서 정현파나 비대칭 삼각형의 lifter가 약 10% 정도의 향상을 보이고 있다. 그리고, closed test의 경우에는 비대칭 삼각형과 정현파 lifter가 비슷한 인식을 보이고 있다. 최대인식은  $hh=0.9$ 일 때 95%였다. 일반적으로 open test의 경우는 closed test에 비하여서 인식이 떨어지는 것으로 알려져 있다.

표 5-2와 표 5-3을 살펴보면 직사각형과 비대칭 삼각형 lifter의 경우에는 인식이 떨어짐을 알 수 있다. 그러나, 정현파 lifter의 경우에는 인식이 떨어지지 않았다. 한편, closed test에서는 비대칭 삼각형과 정현파 lifter가 비슷한 인식을 보이거나 open test의 경우에는 정현파 lifter가 약 7%정

표 5-2 Closed test.

Lifter	인식률	
직사각형	82.5%	
정현파 hh	0.3	92.5%
	0.5	93.8%
	0.7	93.8%
	0.9	95.0%
비대칭삼각형	93.8%	

표 5 - 3 Open test.

Lifter	인식률
직사각형	78.3%
정현파 $hh=0.5$	95.0%
비대칭삼각형	88.3%

\* 정현파 lifter

$$c(k) = 1 + 12 \cdot hh \cdot \sin(3.14k/12)$$

도 나은 것을 알 수 있었다. 즉, 정현파의 lifter가 화자간의 차이를 잘 제거하고 있다. 그러므로, 표 5 - 2와 표 5 - 3에 의하면 화자독립 인식기의 개발에 있어서 정현파 lifter가 가장 좋은 lifter임을 알 수 있다.

## VI. 결 론

본 논문에서는 패턴 정합의 방법에 의거하여 한국어 고립단어 인식 실험을 숫자음을 대상으로하여 수행하였다. 음성신호의 파라미터를 추출하는 방법 으로서는 선형예측에 의한 LPC cepstrum을 사용하였다. 화자 독립인식에 있어서 선형예측방법의 문제 및 그 lifter를 통한 해결방안을 인식실험을 통하여 검토하였다. 또한 기준 패턴을 구하기 위한 방 으로서 집단화에 대하여 UWA방법과 K-iteration을

변형시킨 KMA 방법을 제시 실험 검토하였다.

위와 같은 실험에 의하여 cepstrum 영역에서 정현 파 lifter가 화자의 개인적 성질을 가장 잘 제거함을 알 수 있었고 집단화의 방법으로는 KMA 방법과 같은 평균 기법이 필요함을 알 수 있었다. 또한 인식 단어들을 그 특징에 따라 여러 군으로 묶음으로 인식 시에 비교어휘를 줄일 수 있었다.

## 참 고 문 헌

1. Y. Tohkur, "A weighted cepstral distance measure for speech recognition," ICASSP-86 Proceeding, pp.761-764.
2. B.H. Jung and L.R. Rabiner, "On the use of bandpass liftering in speech recognition," ICASSP-86 Proceeding, pp.765-768.
3. L.R. Rabiner and J.G. Wilson, "Considerations in applying clustering techniques to speaker independent word recognition," IEEE Trans. on ASSP, Vol. ASSP-27, No.2, April 1979.
4. S.E. Levinson and L.R. Rabiner, "interactive clustering techniques for selecting speaker-independent reference template for isolated word recognition," IEEE Trans. on ASSP, Vol. ASSP-27, No.2, April 1979.
5. 김진영, 성평모, "Level crossing과 DPCM을 이용한 유성음/무성음/목음 분류", 전자공학회 포함학술대회 논문집, Vol. II, pp1615-1618, 4월 1986.