# LPC cepstrum과 집단화를 이용한 한국어 고립단어 인식에 관한 연구

김 진영, 성 굉모
서울대학교 공과 대학원 전자공학과

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

Jin Young, Kim  Keong Mo, Sung
Dept. of Electronics Eng. Seoul Nat'l Univ.

## 요 약

본 논문은 화자독립 고립단어 인식에 있어서 LP 모델의 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 방법으로서는 UWA 방법과 K-iteration 방법을 변형시킨 KMA 방법을 제시 비교하였다. 인식 실험 결과 정현파 lifter와 KMA의 집단화 방법을 사용하였을 때 95%의 최고 인식률을 보였다.

## I. 서 론

인간의 음성신호 처리기술은 디지탈 콤퓨터의 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 장애 요소가 되어 왔다[1]. 따라서 LP계수에 대해 변형이 요구되는데 LP계수를 각각 독립적으로 다룰 수가 없음으로 각 계수가 독립적이 되도록 변화시켜야 한다. 이를 위해 각 계수가 독립적인 cepstrum 영역으로의 변환을 사용 조직을 가하게 된다. 한편, 기준패턴을 구하기 위해 집단화의 기법이 요구되는데, Rabiner등에 의해 여러가지가 제안되어 왔다[3,4]. 본 논문에서는 Rabiner가 제안한 UWA 방법과 K-iteration 방법에 평균화 기법을 도입한 KMA 방법을 제시하고 비교하고자 한다.

## II. LPC cepstrum

cepstrum분석은 본래 지진파 연구에 이용되었던 방법으로 그림2-1과 같이 나타낼 수 있다. 입력된 음성 신호의 전력 밀도 스펙트럼을 P(w)라고 할 때 다음과 같이 표시될 수 있다.

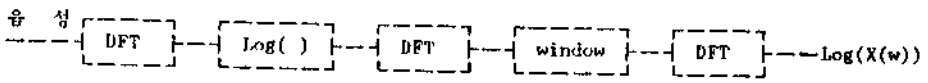



그림2-1. cepstrum 분석

$$P(\omega) = |G(\omega)|^2 |H(\omega)|^2 = |V(\omega)|^2 |E(\omega)|^2 \qquad (1)$$

단, S(w) : 신호의 스펙트럼

G(w) : 음성의 발생 신호

V(w) : 이상적인 음성의 발생 신호

H(w),E(w) : 스펙트럼의 envelope

여기서 quefrency가 제한된 cepstrum C(t)는 다음과 같이 정의 된다.

$$C(t) = i(t) \cdot C'(t) \qquad (2)$$

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

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

여기서 C(t)를 이용하여 분석하는 것을 cepstrum분석 이라고 한다.

1. LPC cepstrum

위에서 cepstrum을 구하기 위하여서 입력 신호의 전력 밀도 스펙트럼을 이용하였다. 그런데 스펙트럼 추정 이론에 의하면 스펙트럼은 선형예측(LP)에 의하여 얻을 수가 있다. LPC에 의하여 구하여진 스펙트럼을 sqrt(a)/A(z)이라고 하면

$$S(\omega) = \sqrt{a}/A(z) \qquad (3)$$

단, $A(z) = \sum_{i=0}^{M} a_i z^{-i}$

M :LPC의 차수

따라서 P(W)대신에 위의 스펙트럼 근사를 이용하여 cepstrum을 구하여 보면

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

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

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

위와같이 cepstrum을 LPC의 스펙트럼을 이용하여 구하는 것을 LPC cepstrum이라고 한다.

2. LPC( LPC cepstrum )의 문제점과 해결

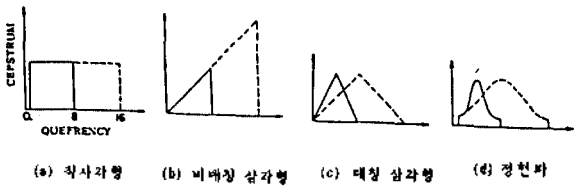
음성 신호의 LP 모델은 pitch와 같은 음성 발생 요소에 어느 정도 둔감하기는 하나 전혀 상관성이 없는 것은 아니다. 또한 all-pole 모델이라는 제약 조건 때문에 음성인식에 바람직하지 못한 스펙트럼 성분을 만들어 낸다[1]. 예를 들어 zero가 없음으로 스펙트럼 영역에서 각 첨두치들의 폭( band width )이 매우 작게 되어 그 중심 주파수가 조금만 바뀌어도 매우 큰 차이를 가져오게 된다. 즉 각 개인의 pitch에 따라 영향을 받게 된다. 따라서 변형을 가해야 하는데 LPC의 계수는 조작이 어려움있다. 왜냐하면 각 계수가 독립적이지 못하기 때문이다. 따라서 앞에서 소개한 cepstrum 변환을 하게되는 것이다. quefrency 영역에 보면 저주파 성분의 pitch가 높은 quefrency에 영향을 주게 된다. 한편 음성이 발생되는 과정에서 음파가 성도를 따라서 진행하게 되는데 이때 고주파 성분이 저주파 성분보다 많은 변화를 겪게 된다. 그런데 각 개인의 성도는 사람마다 다름으로 고주파 성분의 변화는 저주파보다 개인에 따른 변화가 크다. 이러한 영향은 낮은 quefrency 성분에서 나타난다. 따라서 cepstrum의 영역에서 lifter를 사용해야 할 필요성이 있다. lifter로는 여러가지가 있겠으나 그림2-2와 같은 것들이 제안되어 있다[1]. 그림2-2와 같은 lifter를 사용하여 음성인식을 하는 시스템은 그림2-3과 같이 나타낼 수 있다.

그림2-4는 '구'의 '누'에 해당하는 부분의 스펙트럼, 8차의 LPC 모델 스펙트럼, 9, 12, 15차의 cepstrum 스펙트럼 그리고 그것의 lifter를 거친 후의 스펙트럼이다. 여기서 사용한 lifter는 정현파형으로 다음과 같다.

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

단, L : cepstrum 차수

h : 0.5*L




(a) 직사각형 (b) 비대칭 삼각형 (c) 대칭 삼각형 (d) 정현파

그림2-2. lifter의 예

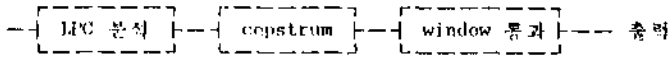



그림2-3. lifter를 사용한 분석

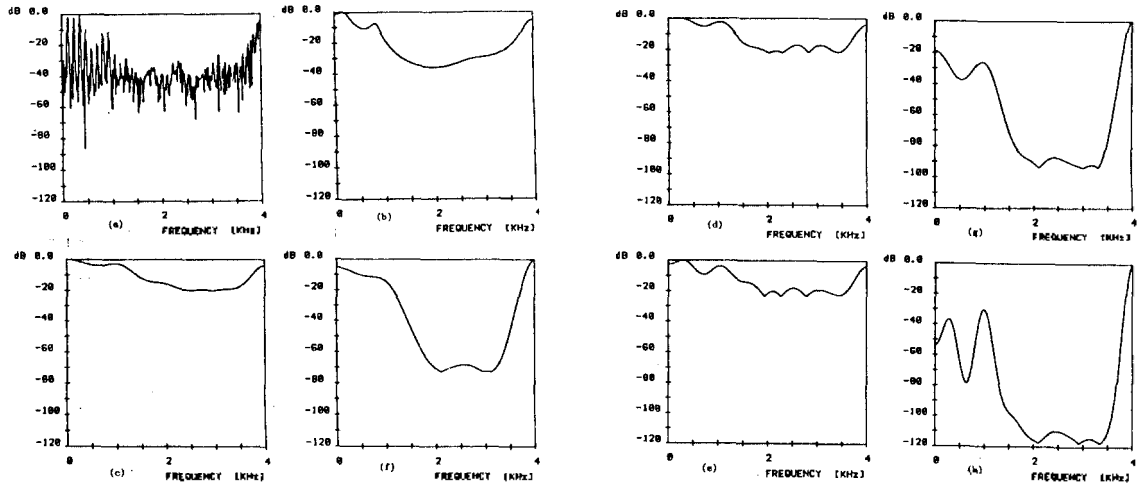



그림2-4. '누'의 스펙트럼 (a) DFT 스펙트럼 (b) LPC 9차의 스펙트럼

cepstrum 스펙트럼 (c) 9차 (d) 12차 (e) 15차

liftered 스펙트럼 (f) 9차 (g) 12차 (g) 12차

## III. 집단화의 기법

패턴 정합의 방법을 이용한 고립단어 인식(특히, 화자독립 )에 있어서 중요한 문제 중의 하나는 인식하고자하는 단어들의 기준 패턴을 만들어 내는 것이다. 이를 위해 따라서 집단화의 기법을 사용하여 가장 최적의 기준 패턴을 만들어야 할 것이다. 지금까지 그 기법으로서 여러가지가 제안되어 왔다. 본 논문에서는 그 중 간단한 방법인 UWA와 K-means iteration의 방법을 변형시킨 KMA( K-means with average )을 제시하고 비교하고자 한다.

### 1. UWA( Unsupervised clustering without averaging )

이 방법은 집단의 결정 threshold값만 주어지면 자연적으로 집단의 갯수와 집단의 원소가 정하여지는 방법이다. 여기서 두 패턴 사이의 거리 즉 Xi와 Xj의 거리는 DTW 방법을 이용하여 구하여진다.

$$d_{ij} = \delta(x_i, x_j) = \frac{1}{N_i} \sum_{k=1}^{N_i} d(i(k), j(k)) \qquad (7)$$

집단들의 집합을 w1,w2,...wj라고 할 때 집단화를 하고자 하는 패턴 중 w1,w2,..,wj에 포함되지 않은 패턴들의 집합 는 다음과 같이 정의된다.

$$\Omega'_{j+1} = \Omega - \Omega'_{i=1} w_j = \Omega'_j - w_j \qquad (8)$$

$$= \{x'_1, x'_2, \cdots, x'_{n(j)}\}$$

단, $\Omega$ : 전체 집합

$x_i'$ : $\Omega_{j+1}'$ 의 요소

$a(j)$ : 처음 $j$개의 집단형성후에 포함되지 않은 수

그러면 UWA 알고리듬은 다음과 같다.

1) 초기화 $j=0$

2) $\Omega_{j+1}'$ 관찰 집합 중 minmax 중심( center )를 구한다. 이 minmax 중심을 $s_{j+1}$라고 한다.

$$\text{minmax} : s_{j+1} = x_{i_0}' \ni \max_j \delta(x_{i_0}', x_j') \le \min_i \max_j \delta(x_i', x_j')$$

3) $w_{j+1}^{(k)}$의 초기 선택(k=0)은 다음과 같다. (9)

$$w_{j+1}^{(k)} = \cup_{i \in \Omega_{j+1}'} x_i' \ni \delta(s_{j+1}, x_i') \le T \quad (10)$$

4) $w_{j+1}^{(k)}$의 minmax 중심 $s_{j+1}$를 선택한다.

5) k를 증가 시키고 다시 $w_{j+1}^{(k)}$의 원소를 결정한다. $w_{j+1}^{(k)} = w_{j+1}^{(k-1)}$이거나 k>KMAX(max iteration)이면 완료된다. 그렇지 않으면 j를 증가시키고 $\Omega_{j+1}$를 결정한다. $\Omega_{j+1}'$ 의 집합이 공집합이 아니면 2)로 간다. 위의 check에 걸리지 않으면 stop로 가라.

2. KMA( K-means iteration with average )

K-means iteration 알고리듬은 특정 수의 집단을 찾는데 유용한 자동화된 반복 과정이다. 이 반복과정은 분류( classification ), 집단 중심의 계산, 수렴 테스트의 세가지 과정으로 이루어져 있다. 만약 M개의 집단을 찾고자 한다면 초기 집단의 중심으로서 M개의 임의의 패턴을 지정해야 한다. 대부분의 음성인식 시스템에서는 집단의 중심을 min max 중심으로 찾으나 여기서는 DTW에 의한 평균 기법을 도입하여, 변형된 KMA 방법을 제시하고자 한다. 그 방법은 아래와 같다.

1) 초기화 $X_i^{(0)} = x_i, \quad 1 \le i \le M$ (11)

2) nearest neighbor 법칙에 의하여 패턴들을 M개의 집단에 귀속 시킨다.

$$x_j \in w_i \quad iff \quad \delta(x_i, x_j^{(i)}) \le \delta(x_j, x_k^{(k)}), \quad 1 \le k \le M \quad (12)$$

여기서, j는 모든 패턴들의 갯수이다.

3) 에 대하여서 DTW를 이용 평균을 취한다.

4) 2),3)을 행한다. 새로운 중심과 과의 중심이 같거나 iteration 수가 주어진 값을 넘으면 완료시킨다.

평균 기법은 다음과 같은 과정을 통한다. 즉 두 패턴 xi와xj를 평균하려면 DTW를 통하여 최적의 path의 werping 함수를 찾는다.

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

$$m = i(k), \quad n = j(k), \quad k = 1, 2, \cdots, K \quad (13)$$

이 된다. 이것을 기준 프레임 수로 만들기 위해서는 interpolation을 하면 된다. 다른 패턴에 대해 위와같은 과정을 반복한다.

## IV. 실험및 검토

### 1. 실험 데이타의 제집

논문에서는 인식 대상 어휘로서 음성 다이얼링( Voice-dialing )을 위한 한국어 숫자음을 사용하였다. 데이타의 수집은 4명의 화자가 각각 3번씩 발음하여 총 120개의 숫자음을 얻도록 하였다.

### 2. UWA와 KMA의 비교

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

여기서는 단순히 두 집단화 알고리듬을 비교하고자 하였으므로 학습패턴들과 시험 패턴들을 같게 하였다. 표4-1은 기준 패턴이 하나인 경우와 둘인 경우의 인식률을 나타낸 것이다. 표4-1에 의하면 KMA 방법이 UWA방법에 비하여서 11,5% 정도 인식률이 나은 것을 알수 있다. 인식 실험시 이 원인을 살펴본 결과 UWA 방법에 있어서의 오인식은 화자에대한 의존성 때문에 생긴 것으로 보인다. 즉, UWA 의 방법은 minmax에 의하여 기준 패턴을 구하는데 이것이 어느 특정화자의 발음한 패턴이기 때문이다.

표4-1. UWA와KMA의 비교(DTW window=6)

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

2. Liftering에 의한 인식 비교

인식실험의 결과를 표4-2과 표4-3에 보였다.

표4-2에 의하면 직사각형의 lifter의 경우에 비하여서 정현파나 비대칭 삼각형의 lifter가 약 10%정도의 향상을 보이고 있다. 그리고, closed test의 경우에는 비대칭형 삼각형과 정현파 lifter가 비슷한 인식률을 보이고 있다. 최대 인식률은 hh=0.9일때 95% 였다. 일반적으로 open test의 경우는 closed test에 비하여서 인식률이 떨어지는 것으로 알려져 있다. 표4-2와 표4-3을 살펴보면 직사각형과 비대칭 삼각형 lifter의 경우에는 인식률이 떨어짐을 알 수 있다. 그러나 ,정현파 lifter의 경우에는 인식률이 떨어지지 않았다. 한편, closed test에서는 비대칭 삼각형과 정현파 lifter가 비슷한 인식률을 보이나 open test의 경우 에서는 정현파 lifter가 약 7%정도 나은 것을 알 수 있었다. 즉, 정현파의 lifter가 화자간의 차이를 잘 제거하고 있다. 그러므로, 표4-2 와 표4-3에 의하면 화자독립 인식기의 개발에 있어서 정현파 lifter가 가장 좋은 lifter임을 알 수 있다.

표4-2. 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% |

표4-3. Open test

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

* 정현파 lifter
c(k)=1+12hh*sin(3.14k/12)

V. 결 론

본 논문에서는 패턴 정합의 방법에 의거하여 한국어 고립단어 인식 실험을 숫자음을 대상으로 하여 수행하였다. 음성신호의 파라미터를 추출하는 방법으로서는 선형예측에 의한 LPC cepstrum을 사용하였다. 화자 독립 인식에 있어서 선형 예측 방법의 문제및 그 lifter를 통한 해결 방안을 인식 실험을 통하여 검토하였다. 또한 기준 패턴을 구하기 위한 방법으로서 집단화에 대하여 UWA 방법과 K-iteration을 변형시킨 KMA 방법을 제시 실험 검토 하였다.

위와 같은 실험에 의하여 cepstrum 영역에서 정현파 lifter가 화자의 개인적 성질을 가장 잘 제거함을 알 수 있었고 집단화의 방법으로서는 KMA 방법과 같은 평균 기법이 필요함을 알 수 있었다.

참고문헌


[1] B.H.Jung and L.R.Rabiner,"On the use of bandpass liftering in speech recognition," ICASSP.86 Proceeding

[2] L.R.Rabiner and A.E.Rosenberg,"considerations in dynamic time warping algorithms for discrete words recognition," IEEE Trans. on ASSP, Vol. ASSP-26, December 1978

[3] L.R.Rabiner and J.G.Wilson,"considerations in applying clustering techniques to speaker independent word recognition ," J.A.S.A., Vol.66, September 1979

[4] 김 계곡," 집단화를 이용한 한국어 숫자음의 기준패턴 설정에 관한 연구," 한국 음향학회지 제5권 2호, 1986