# PCA 기반 파라메타를 이용한 숫자음 인식

박경훈, 표창수, 김창근, 허강인
동아대학교 전자공학과

# The Recognition of Korean Syllables using Parameter Based on Principal Component Analysis

Kyung Hoon Park, Chang Soo Pyo, Chang Keun Kim, Kang in Hur
Dept. of Electronics, Dong-A University
E-mail : kihur@mail.donga.ac.kr

## 요 약

본 논문에서는 음성 특징추출의 한 방법으로서 기존의 방법들과는 달리 음성의 통계적인 특성들을 고려하여, 입력 공간내에서 변동량이 가장 많은 방향으로 주축을 발견한 다음 그 정보를 이용하여 데이터의 중복성을 제거하는 주성분 해석(PCA:Principal Component Analysis)기법을 사용하여 음성의 특징을 추출하는 방법을 제안한다. 본 논문의 숫자음 인식실험 결과와 비교하기 위하여 기존의 음성특징 파라메타인 Mel-Cepstrum과 비교하였을 때, 0.5%의 인식률 차이가 있었으나, 음성특징 추출시 기존의 파라메타에 비하여 비교적 짧은 시간에 구해지는 점과 데이터의 통계적 특성을 이용한 최적의 기저벡터를 이용한다면 단어나 문장 인식시에 보다 나은 인식률을 얻으리라 사료된다.

## ABSTRACT

The new method of feature extraction is proposed, considering the statistic feature of human voice, unlike the conventional methods of voice extraction. PCA(Principal Component Analysis) is applied to this new method. PCA removes the repeating of data after finding the axis direction which has the greatest variance in input dimension. Then the new method is applied to real voice recognition to assess performance. When results of the number recognition in this paper and the conventional Mel-Cepstrum of voice feature parameter are compared, there is 0.5% difference of recognition rate. Better recognition rate is expected than word or sentence recognition in that less convergence time than the conventional method in extracting voice feature. Also, Better recognition rate is expected when the optimum vector is used by statistic feature of data.

## Ⅰ. 서 론

음성은 인간의 가장 자연스러운 의사소통의 수단이며 이러한 음성에 의한 인간-기계의 인터페이스는 속도가 빠르고 특별한 훈련이 없어도 이루어진다. 또한 컴퓨터 및 정보통신 기술의 급속한 발전으로 음성인식 기술은 중요한 연구과제가 되어지고, 오늘날까지 음성인식의 성능을 향상시키기 위해서 많은 연구가 되어지고 있다. 특히 음성인식기 부분은 현재까지도 좋은 성능을 보여주는 인식기가 많이 사용되고 있다. 그러나, 패턴인식이나 분류능력에는 한계가 있으며 결국에는 인식 성능 향상을 위해서 가장 중요한 과제는 인식기의 입력으로 들어가는 음성의 특징을 효율적으로 선택하는 것이라 할 수 있다. 즉, 각각의 패턴들의 특징들을 가장 잘 반영하는 특징을 추출함으로써 그 특징들을 인식기의 입력으로 사용하

여 인식 성능을 더욱 향상 시킬 수 있는 것이다. 이렇게 추출된 음성특징 파라메타는 음성신호의 특징을 나타내고 있으므로 음성정보의 압축효과를 얻을 수 있을 뿐만 아니라, 음성인식을 위해 필수적이다.[1] 본 논문에서는 음성의 통계적인 특성들을 고려하여, 입력 공간내에서 변동량이 가장 많은 방향으로 주축을 발견한 다음 그 정보를 이용하여 데이터의 중복성을 제거하는 주성분 해석(PCA:Principal Component Analysis)기법을 사용하여 음성의 특징을 추출하는 방법을 제안하고, 실제로 음성인식에 적용함으로써 인식성능을 평가해 보았다. 본 논문에서는 PCA가 음성 인식에서의 특징 추출 방법에 적용가능한지에 대해서 연구의 초점을 맞추고자 한다. 그림 1은 PCA를 이용한 특징추출 알고리즘의 전체 블록도를 도식화한 것인데, 입력음성에 대하여 주성분해석기법(PCA)을 적용하여 기저벡터를 구한 후 기저벡터 계수를 고려하여 그 기저벡터와 입력음성과의 상관관계에 의하여 음성의 특징을 추출하게 된다.

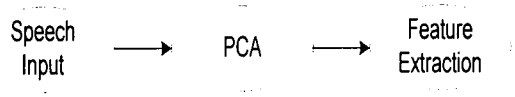


그림1. 전체 블록도

본 논문의 전체 구성은 다음과 같다. 2장에서는 본 논문에서 제안한 알고리즘인 주성분해석(PCA)기법에 대해서 설명을 하고 3장에서는 PCA를 적용하여 생성된 기저벡터를 음성 인식기의 입력으로 사용하여 인식 성능을 알아본다. 마지막으로 4장에서는 실험결과에 대해 논의하고 끝으로 결론을 짓고자 한다.

## II. 배경이론

### 2. 주성분해석(PCA:Principal Component Analysis)기법

음성인식에서 가장 중요한 문제는 각각의 음성들의 특성을 가장 잘 반영하는 특징을 추출(Feature Extraction)하는 것이다. 이러한 특징추출 문제는 실제 데이터에서는 본질적이고 중요한 문제이다. PCA(Principal Component Analysis)는 입력의 선형성과 특성 식별을 이용하여 데이터의 차원을 축소하며 특징을 추출하는 방법 중 하나로, 패턴인식에서 Karhunen-Lőeve 변환으로 잘 알려져 있다.[2][3]

zero-mean 특성이 있는 $n$차원 신호 $x$에 대한 단위 벡터를 $v_i$라고 하면, $i=1,2,\cdots n$인 $n$개의 가능한 투영 $p_i$는 식 (1)과 같다.

$$p_i = x \cdot v_i = x^T v_i = v_i^T x \tag{1}$$

$p_i$는 $x$를 단위 벡터 같은 차원의 $v_i$ 영역으로 투영시킨 것으로, 각각의 단위 벡터에 대한 특성을 나타내는 스칼라 $p_i$를 벡터로 표시하면 식 (2)와 같다.

$$\begin{aligned} p &= \{p_1, p_2, \cdots p_n\} \\ &= \left[x^T v_1, x^T v_2, \cdots, x^T v_n\right]^T \\ &= V^T x \end{aligned} \tag{2}$$

식 (2)에서 $p$와 단위 벡터 $v_i$를 이용하여 입력신호 $x$를 복원하는 방법은 식 (3)과 같으며, 이는 입력신호를 단위벡터와 그에 대한 투영 값에 의하여 나타낸다.

$$\begin{aligned} x &= V^T p \\ &= \sum_{i=1}^{n} p_i v_i \end{aligned} \tag{3}$$

이상은 입력신호 $x$를 단위벡터 $v_i$를 이용하여 표현한 것으로, 차원에 대한 축소는 없었고 단지 좌표변환만 있었다. 다음은 $v_i$와 $p$의 차원을 줄여, 좀더 작은 차수로 $x$를 근사하는 방법을 설명한다. 식 (3)에서 $i$를 $n$에서 $m$(단, $m<n$)으로 축소한 $x$를 $\hat{x}$이라고 하면, $\hat{x}$는 식 (4)와 같이 나타낼 수 있다.

$$\hat{x} = \sum_{i=1}^{m} a_i u_i \tag{4}$$

여기서 $a_i$와 $u_i$는 $x$의 공분산 행렬(covariance matrix) $R$의 고유치와 고유벡터로 중 일부로, $R$를 구하는 방법은 식(5)에, 그에 대한 고유벡터와 고유치를 구하는 방법은 식 (6)에 나타내었다. 식 (6)에서 $n$개의 $R$의 고유벡터 $\lambda$를 내림차순 정렬한 것 중 상위의 $m$개를 취한 것을 $a$로, 이에 순서에 따른 $u_i$를 $\overline{u_i}$로 나타내면 식 (7)과 같다.

$$R = \frac{1}{n} \sum_{i=1}^{n} (x x^T) \tag{5}$$

$$Ru_i = \lambda u_i \quad , i = 1, 2, \cdots n \tag{6}$$

$$R\overline{u_i} = a\overline{u_i} \quad , i = 1, 2, \cdots m \tag{7}$$

원래의 $x$와 근사한 $\hat{x}$의 오차를 $e$라고 하면 그 크기는 식 (8)과 같다.

$$e = \sum_{m+1=n}^{n} a_i \overline{u_i} \tag{8}$$

## III. 인식실험

3. 음성데이타 및 분석조건

본 논문에서 사용한 음성데이터는 ETRI의 샘돌이 음성 데이터 중에서 "공,일,이,삼,사,오,육,칠,팔,구" 10개의 음성을 사용하였다. 이는 남성화자 20명이 10개 숫자음을 4회 발성한 총 800개의 데이터 중에서 PCA를 이용하여 기저벡터를 구하기 위해서 20명이 3회 발성한 600개의 데이터를 사용하였고, 나머지 200개의 데이터를 인식기의 테스트용으로 사용하였다. 음성 신호는 16kHz 표본화 비율에서 16bit로 양자화 하였다. 음성 부분(Speech Segment)은 60개의 표본개수를 갖고 있으며 이는 약 3.75ms의 시간 구간에 해당된다. 본 논문에서는 그림 2에 있는 left-to-right형 모델인 연속출력분포 HMM을 사용하였고 각 숫자음 모델은 16상태로 표현되었다.[4][5]

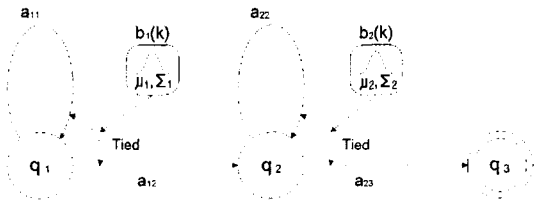


그림2. 연속 출력분포 HMM

실험에서 60개의 표본개수를 가진 모든 입력음성에 대하여 PCA(Principal Component Analysis) 알고리즘을 적용하면 60×60개의 기저벡터가 생성된다. 이 생성된 기저벡터는 기저벡터계수에 의해 중요도가 높은 순서로 나열되어 있다. 그림 3은 각 기저벡터에 대한 중요도를 나타내는 기저벡터계수들을 나타낸 것이다.

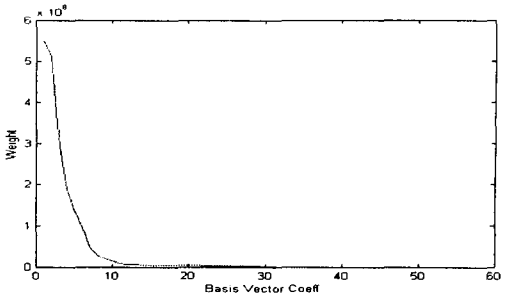


그림3. 기저벡터계수

그림3에서 기저벡터계수가 10개에서 급격하게 떨어짐을 알수가 있다.
또한 그림4는 1번째, 2번째, 10번째, 20번째 기저벡터에 대한 주파수 특성을 보여주고 있다.

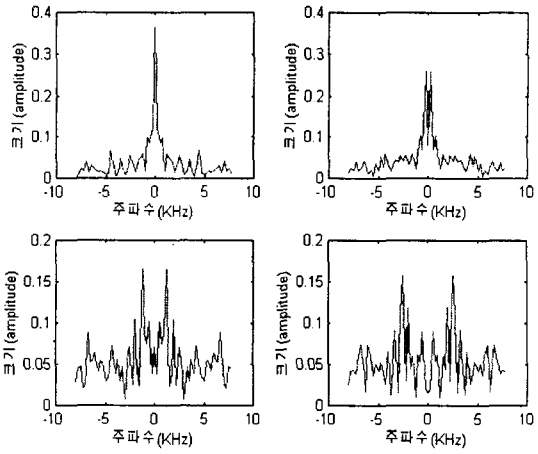


그림4. 각 기저벡터에 대한 주파수특성

그림5는 PCA로 얻어진 60개의 기저벡터들의 중심주파수(Center Frequency)를 순서대로 나타내고 있다. 기저 벡터들의 중심주파수들은 250Hz~8kHz범위에서 선형적으로 분포함을 볼 수 있고, 거의 낮은 주파수 성분에서부터 높은 주파수 성분으로 배열되어있음을 볼 수 있는데 이는 인간의 음성신호는 상대적으로 낮은 주파수 성분에 더 많은 에너지를 가지고 있다는 사실에서 기안한다. 위의 사실로부터 PCA를 이용하여 음성인식에 적용하기 위해서는 기저벡터들로부터 중요한 특징벡터들을 선택할 수 있다.

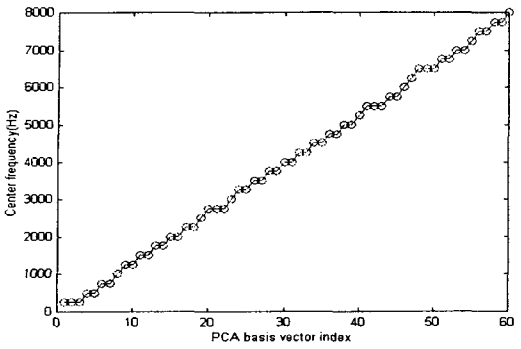


그림5. 기저벡터들의 중심주파수

PCA는 입력차원에 해당하는 독립 요소들을 찾아내고, 이 중에 불필요하거나 중복되어 있는 것들이 발생할 수 있다. 이러한 불필요함이나 중복됨을 줄여주기 위해서 본 논문에서는 기저벡터계수의 가중치의 변화량을 고려하여 기저벡터의 차원을 선택하였다. 이 선택된 기저벡터는 입력음성과의 상관관계에 의하여 차원이 줄어들게 되어서 인식기의 입력특징 파라메타화 되는 것이다.

## Ⅳ. 실험결과 및 결론

표1은 HMM의 상태수에 따른 Mel-Cepstrum음성특징과 제안된 방법으로 추출된 음성특징을 선택된 기저벡터 개수를 변화시키면서 실험한 인식결과를 비교해서 보여준다. 기저벡터계수의 기여도(가중치)에 따라 15개의 기저벡터를 선택했을 때 제안된 음성특징의 인식률이 98.5%였다.

| | 12 State | 15 State | 16 State |
|---|---|---|---|
| Mel-Cepstrum | 99.0 | | |
| 10 Basis vector | 97.0 | 97.0 | 96.5 |
| 15 Basis vector | 98.0 | 98.0 | 98.5 |
| 20 Basis vector | 97.5 | 97.5 | 97.5 |

표1. 기저벡터 개수에 따른 인식률변화 (%)

본 논문에서는 PCA알고리즘으로 추출된 기저벡터를 기저벡터계수를 고려하여 음성인식에 적용해 보았다. 숫자음 인식 결과 기존의 Mel-Cepstrum특징 파라미터에 비해서 0.5%의 인식률 저하가 있었다. 그러나 PCA알고리즘은 입력 데이터의 통계적인 특성을 이용하기 때문에 최적의 기저벡터를 이용한다면 일부분의 기저벡터를 이용하여서도 인간의 음성신호를 효과적으로 부호화 할 수 있는 특징 추출의 한 방법임을 알 수가 있었다. 앞으로 단음절이 아닌 단어나 문장의 음성데이타를 이용한 음성인식실험과 기저벡터를 이용한 음성인식 외에 음성처리 등의 분야에서 많은 연구가 되어져야 할 것이고 비교적 짧은 시간에 의한 파라메타 추출로 인하여 실시간 음성인식에 적용가능성이 충분하다고 사료된다.

## Ⅴ 참고문헌

[1] L.Rabiner and B. Juang. "Fundamentals of Speech Recognition", Prentice Hall, 1993.

[2] T.W.Lee. "Independent Component Analysis-Theory and Applications", Kluwer Academic Publishers, 1998.

[3] J.H.LEE. "On the Efficient Speech Feature Extraction Based on Independent Component Analysis", 1999.

[4] 심장엽, 이영재, 고시영, 이광석, 허강인. "HMM에 의한 연속음성인식 시스템의 구현", 제13회 음성통신 및 신호처리 워크샵 논문집 제13권 1호, pp.325-330, 1996.08.

[5] 표창수, 김창근, 허강인."HMM의 출력확률을 이용한 신경회로망의 성능향상에 관한 연구", 신호처리 시스템 학회 논문지, 제 1권 1호, pp. 1-6, 2000.10