독립성분분석법을 이용한 음성인식기의 성능향상

김창근, 한학용, 허강인
동아대학교 전자공학과

Performance Improvement of Speech Recognition Based on Independent Component Analysis

Chang-Keun Kim, Hag-Yong Han, Kang-In Hur
Dept. of Electronics, Dong-A University
E-mail : ckkim@pattern.donga.ac.kr

요 약

본 논문에서는 신호간의 의존성과 관련성이 최소가 되도록 분리하는 독립성분분석법을 이용하여 입력음성에서 변동량이 많은 방향으로 주축을 찾아 그 정보를 이용하여 데이터의 중복성을 제거한 후 음성특징벡터를 추출하는 방법을 제안한다. 학습하고자하는 음성인식기의 음성에서 독립성분분석법을 이용하여 특징벡터를 추출하고 HMM을 사용하여 기존의 음성특징벡터로 사용되는 mel-cepstrum과 비교하여 학습, 인식실험을 수행하였으며 제안한 방법에서 음성인식성능의 향상을 확인할 수 있었다. 또한, 인식 시 주변여건에 따라 잡음에 의한 인식성능 저하에도 유연히 대처할 수 있음을 알 수 있었다.

Keywords : speech recognition, feature extraction, independent component analysis, basis vectors

Abstract

In this paper, we proposed new method of speech feature extraction using ICA(Independent Component Analysis) which minimized the dependency and correlation among speech signals on purpose to separate each component in the speech signal. ICA removes the repeating of data after finding the axis direction which has the greatest variance in input dimension.

We verified improvement of speech recognition ability with training and recognition experiments when ICA compared with conventional mel-cepstrum features using HMM. Also, we can see that ICA dealt with the situation of recognition ability decline that is caused by environmental noise.

Ⅰ. 서론

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 음성인식기가 많이 연구, 개발되어 지고 있다. 그러나, 패턴인식이나 분류능력에는 한계가 있으며 결국에는 인식 성능 향상을 위해서 가장 중요한 과제는 인식기의 학습이나 인식대상이 되는 입력 음성의 특징을 분류의 측면에서 효율적으로 선택하는 것이라 할 수 있다. 즉, 각각의 패턴들의 특징들을 가장 잘 반영하는 특징을 추출함으로써 그 특징들을 인식기의 입력으로 사용하여 인식 성능을 더욱 향상시킬 수 있는 것이다. 이렇게 추출된 음성특징 파라메타는 음성신호의 특징을 나타내고 있으므로 음성정보의 압축효과를 얻을 수 있을 뿐만 아니라, 음성인식을 위해 필수적이다. 또한, 실시간으로 인간의 음성을 인식할 수 있는 인식기를 개발하는 데에는 적은 양의 정보로 음성신호를 효과적으로 표현하고 처리할 수 있도록 하는 신호의 특징을 추출하는 것이 중요한 연구과제라 할 수 있다.

음성데이터에서 최적의 특징을 추출하기 위해서 최근 신호들의 통계적인 특성으로부터 데이터의 특징을 추출하는 기법이 많이 연구되고 있다.

그 중에서 음성의 통계적인 특성들을 고려하여 입력공

간 내에서 변동량이 가장 많은 방향으로 주축을 발견한 다음, 그 정보를 이용하여 데이터의 중복성을 제거하는 주성분해석(Principal Component Analysis)기법을 사용하여 음성의 특징을 추출하는 방법이 대표적이다.

그러나, 음성신호는 독립적인 고차 통계특성으로 구성되어지므로 통계적으로 독립인 성분을 추출하여 음성의 특징으로 사용할 수 있는 독립성분분석법(Independent Component Analysis)이 관심을 끌고 있다.

그림1은 독립성분분석법을 이용한 특징추출의 블록도를 도식화한 것인데, 입력음성에 대하여 독립성분분석법을 적용하여 기저벡터를 구한 후 기저벡터계수를 고려하여 그 기저벡터와 입력음성과의 상관관계에 의하여 음성의 특징을 추출하게 된다.

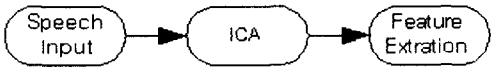

그림 1. 블록도

본 논문의 전체 구성은 다음과 같다. 2장에서는 본 논문에서 제안한 알고리즘인 독립성분분석(ICA)기법에 대해서 설명을 하고 3장에서는 독립성분분석법을 적용하여 생성된 기저벡터를 사용하여 특징을 추출한 다음, 음성인식기의 입력으로 사용하여 인식 성능을 알아본다. 마지막으로 실험결과에 대해 논의하고 결론을 짓고자 한다.

II. 독립성분분석

독립성분분석은 특정신호의 생성단계에서 선형적으로 혼합되어 있는 독립신호원들을 관측데이터에서 분리해내는 방법이다. 임의의 신호는 몇 개의 확률적으로 독립인 신호들에 가중치가 곱해진 다음 혼합되어 생성된 것이라 가정하고 정보이론, 신경회로망의 비교사학습방법, 통계적신호처리, 베이즈 확률이론 등을 이용하여 여러 신호간의 통계적인 의존성을 정의하고, 의존성이 최소가 되는 가중치를 추정하여 임의의 신호가 생성되게 하는 독립신호원을 구할 수 있다.

이러한 방법은 특정 신호로부터 기본 구성 요소를 이루는 독립 성분들을 얻어내는데 적용되어 많은 성공적인 결과를 나타내고 있다. 이러한 독립성분 분석기법을 이용하여 잡음환경에서의 음성신호 인식이나 음성신호 특징추출, 음질개선, 영상정보 코딩, 의료용 신호처리와 칵테일파티문제로 정의되는 여러 가지 신호에서 특정한 신호만 취하는 선택적 주의집중 문제 등, 여러 가지 문제의 해결에 적용하는 연구가 진행되고 있다.

음성 신호는 통계적으로 독립인 고차 신호 특성들로 구성되어 있으며 고차 신호 특성들은 음성 신호의 주파수와 위상 스펙트럼을 나타내는 기저함수와 관계된 필터를 통해서 추출될 수 있다. 음성 신호로부터 독립 고차 신호 특성들을 분리해내기 위한 기저 함수들은 엔트로피 최대화 방법을 통하여 학습시킬 수 있다. 이렇게 학습된 기저 함수들은 음성 신호의 특정 주파수 대역에 민감한 특성을 보인다.

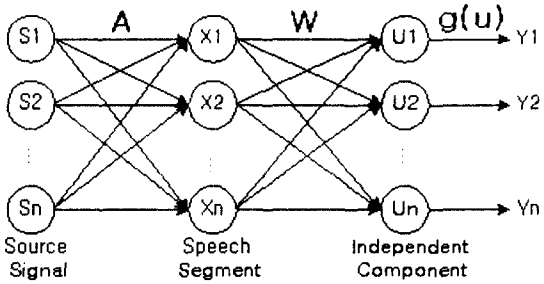

그림 2. 기저벡터학습을 위한 독립성분분석 블록도

그림2는 독립성분분석법을 사용하여 음성특징을 추출하는 전체 블록도이다. 여기서, n개의 평균이 0인 확률변수 $s_1(t), s_2(t), \cdots s_n(t)$로 이루어진 입력신호는 입력벡터 $\boldsymbol{s}(t)=[s_1(t), s_2(t), \cdots s_n(t)]$로 표현할 수 있고 확률적으로 독립이다. 그리고, 각 시점에서 관측되어지는 벡터 $\boldsymbol{x}(t)=[x_1(t), x_2(t), \cdots x_n(t)]$는 다음과 같이 입력벡터 $\boldsymbol{s}(t)$와의 선형결합으로 표현할 수 있다.

$$\boldsymbol{x}(t)=\boldsymbol{A}\boldsymbol{s}(t) \qquad (1)$$

여기서, $\boldsymbol{A}$는 선형혼합행렬이라 한다. 선형혼합행렬을 이용하여 생성되어지는 관측벡터 $\boldsymbol{x}(t)$의 각 성분들은 더 이상 독립적이지 않다. 여기서 우리는 확률적으로 독립인 신호성분을 가지고 있는 $\boldsymbol{s}(t)$를 이용하여 특정신호의 특징벡터로서 사용할 수 있을 것이다.

혼합행렬 $\boldsymbol{A}$에 의하여 벡터 $\boldsymbol{x}(t)$가 관측되어지므로 혼합행렬 $\boldsymbol{A}$나 그 역행렬인 $\boldsymbol{W}$를 구하여 독립소스입력 $\boldsymbol{s}(t)$를 추출할 수 있다. 추정된 독립성분들의 통계적인 독립성은 상호정보량(mutual information)로 정의한다. 관측벡터의 상호정보량은 추정된 독립성분들의 결합엔트로피와 각각의 엔트로피의 차로 계산을 하거나 Kullback-Leibler 발산정리에 의해 다음과 같이 정의한다.

$$I(\boldsymbol{x})=\int p(\boldsymbol{x})\log\frac{p(\boldsymbol{x})}{\prod_{i=1}^{N}p_i(x_i)}d\boldsymbol{x} \qquad (2)$$

상호정보량은 항상 양수이고, 또한 각 성분들이 독립적으로 분리되었을 때는 0 이 된다.

독립성분분석의 목적은 선형변환행렬인 $\boldsymbol{W}$를 상호정보량을 최소화하는 방향으로 학습하여 독립성분벡터 $\boldsymbol{u}(t)$를 구하는 것이다. 이 과정을 다음과 같이 표현할 수 있다.

$$\boldsymbol{u}(t)=\boldsymbol{W}\boldsymbol{x}(t)=\boldsymbol{W}\boldsymbol{A}\boldsymbol{x}(t) \qquad (3)$$

여기서, $u(t)$는 입력벡터 $s(t)$의 추정치이다.

독립성분분석의 학습방법은 결합엔트로피 $H(y)$의 최대화방법을 사용하였고 다음과 같이 표현한다.

$$\Delta W \propto \frac{\partial I(y, x)}{\partial W} = \frac{\partial H(y)}{\partial W} \quad (4)$$

$$\Delta W \propto [\ W^T]^{-1} + (\frac{\partial p(u)}{\partial u} / p(u))\ X^T \quad (5)$$

여기서, $g(u)$는 독립입력신호 $s(t)$의 누적분포함수의 추정인 비선형함수이며, $p(u)$는 관측신호의 확률밀도함수이고 다음과 같이 표현한다.

$$p(u_i) = \frac{\partial y_i}{\partial u_i} = \frac{\partial g(u_i)}{\partial u_i} \quad (6)$$

또한, 빠른 학습을 위해서 W의 역행렬 계산이 필요하지 않은 다음의 수식으로 표현되는 자연감소법(natural gradient)을 이용한다.

$$\Delta W \propto \frac{\partial H(y)}{\partial W}\ W^T W = [I - \varphi(u) u^T] W \quad (7)$$

여기서, $\varphi(u)$는 비용함수라 부른다.

III. 실험 및 고찰

본 논문에서 사용한 음성 데이터는 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 독립성분분석법을 이용하여 기저벡터를 구하기 위해서 20명이 3회 발성한 600개의 데이터를 사용하였고, 나머지 200개의 데이터를 인식기의 테스트용으로 사용하였다. 음성 신호는 16kHz 표본화 비율에서 16bit로 양자화 하였다. 음성부분(Speech Segment)은 60개의 표본개수를 갖고 있으며 이는 약 3.75ms의 시간 구간에 해당된다. 본 논문에서는 그림3에 있는 left-to-right형 모델인 연속출력분포 HMM을 사용하였고 각 숫자음 모델은 16상태로 표현되었다.

그림4는 독립성분분석법에 의한 음성인식실험에 대한 전체 과정에 대한 블록도이다.

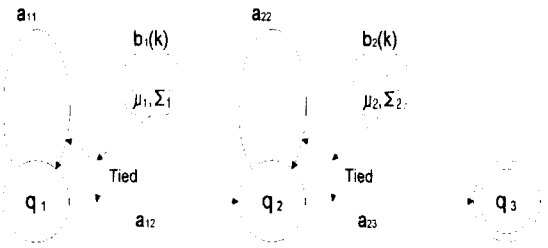

그림 3. 연속출력분포 HMM

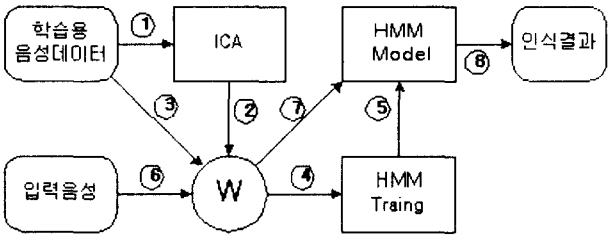

그림 4. ICA에 의한 전체 음성인식실험 블록도

실험에서는 60개의 표본개수를 가진 모든 입력음성에 대하여 주성분분석법(PCA)과 독립성분분석법(ICA)을 적용하여 기저벡터를 학습하였으며 최종적으로 주성분분석법과 독립성분분석법을 이용한 경우에 같은 학습조건으로 60×60개의 기저벡터를 생성하였다. 이렇게 생성된 기저벡터는 주성분분석법인 경우에는 기저벡터계수에 의해 중요도가 높은 순서로 나열되어 지지만 독립성분분석인 경우에는 중요도의 순서로 정렬되지 않기 때문에 중요도를 다시 계산하여야 한다.

그림 4는 독립성분분석법으로 추출한 기저벡터의 L2-norm을 계산하여 기저벡터들의 중요도를 다시 순서대로 정렬한 그림이다. 본 실험에서는 주성분분석법과 독립성분분석법으로 추출한 각 기저벡터에 대하여 중요도를 기준으로 각각 10개, 15개, 20개를 선택하여 인식실험을 수행하였다.

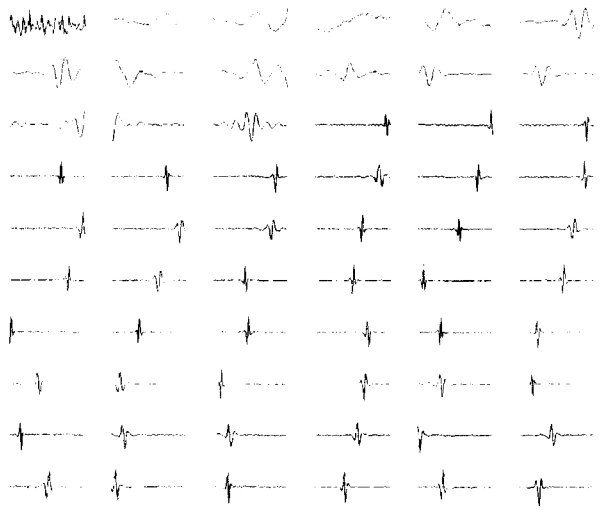

그림 5. 독립성분분석법에 의한 기저벡터

또한, 그림6과 그림7은 1번째, 2번째, 10번째, 20번째 기저벡터에 대한 주파수 특성을 보여주고 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 기초한다 . 위의 사실로부터 주성분분석법과 독립성분 분석법을 이용하여 기저벡터들로부터 중요한 특징벡터들을 추출하여 음성인식에 적용할 수 있음을 알 수 있다.

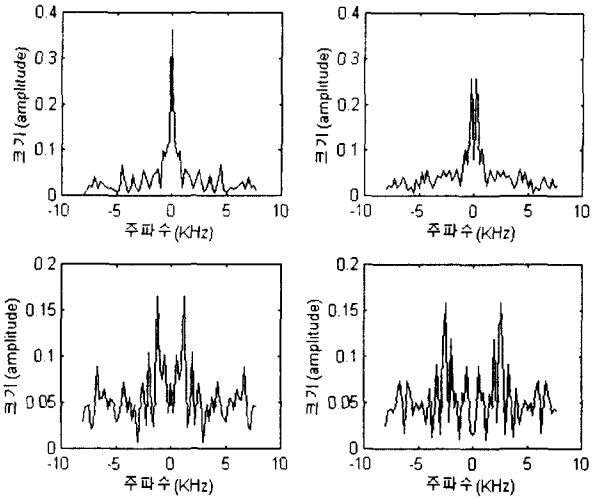

그림6. PCA에 의한 각 기저벡터에 대한 주파수특성

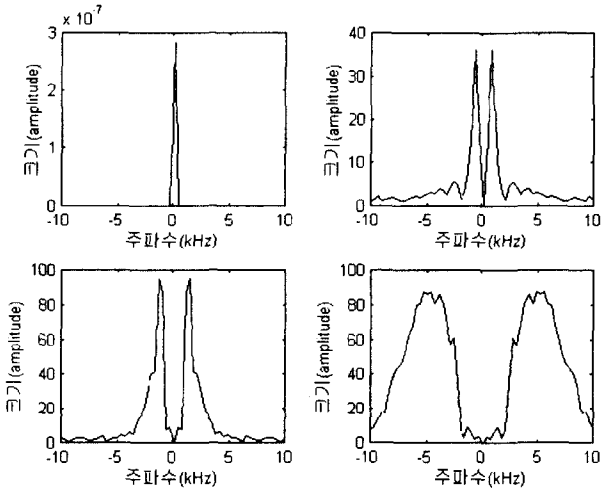

그림 7. ICA에 의한 각 기저벡터에 대한 주파수특성

독립성분분석법으로 관측데이터의 입력차원에 해당하는 독립 요소들을 찾아내고, 이 중에 불필요하거나 중복되어있는 것들이 발생할 수 있음을 알 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. 결론

본 논문에서는 독립성분분석법으로 추출된 기저벡터를 기저벡터계수를 고려하여 음성인식에 적용해 보았으며 기존의 방법인 주성분분석방법과 비교하여 보았다. 숫자음 인식 결과 기존의 mel-cepstrum특징 파라미터에 비해서 주성분분석법의 경우에는 0.5%의 인식률 저하가 있었지만 독립성분분석의 경우에는 기저벡터의 개수와 학습방법에 의해서 차이가 있지만 평균 1~2%의 인식률향상이 있었다. 본 실험에서 사용한 음성데이터의 부족으로 인하여 적절한 최적의 기저벡터를 구하지는 못하였지만 실시간 음성인식장치를 위한 소규모의 음성데이터로서 최적의 결과를 얻을 수 있는 방법임을 확인할 수 있었고, 독립성분분석법은 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 연속음성데이터를 이용한 인식실험과 기저벡터를 이용한 음성인식 외에 음성처리 등의 분야에서 많은 연구가 되어져야 할 것이고 비교적 짧은 시간에 의한 파라메타 추출로 인하여 실시간 음성인식에 적용가능성이 충분하다고 사료된다.

참고문헌

[1] Bell A.J. and Sejnowski T.J. : "An Information-Maximization Approach to Blind Separation and Blind Deconvolution", *Neural Computation*, 1995, 7, 1129-1159

[2] Bell A.J. and Sejnowski T.J. : "The Independent Component of natural scenes are edge filters", *Vision research*, 1997, vol37, (23), 3327-3338

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

[4] Amari S., Cichochi A., and Yang H. : "A new learning algorithm for blind signal separation", *Advances in Neural Information Processing Systems*, 1996, 8, pp.757-763

[5] Simon Haykin : "Neural Networks - A Comprehensive Foundation", *Prentice Hall*, 1999

[6] 박경훈,표창수,김창근,허강인 : "PCA기반 파라메타를 이용한 숫자음 인식",*한국신호처리시스템학회 추계학술대회논문집*, 2000, 제1권2호, 181-184

[7] 표창수,김창근,허강인 : "HMM의 출력확률을 이용한 신경회로망의 성능향상에 관한 연구", *한국신호처리시스템학회 논문집*, 2000, 제1권1-1호, 1-6