

논문 2004-41SP-3-21

SVM 음성인식기 구현을 위한 강인한 특징 파라미터

(Robust Feature Parameter for Implementation of Speech Recognizer Using Support Vector Machines)

김 창 근*, 박 정 원**, 허 강 인*

(Chang Keun Kim, Jeong Won Park, and Kang In Hur)

요 약

본 논문은 두 가지 비교 실험을 통하여 효과적 음성인식 시스템을 제안한다. 분별적 이진 패턴 분류기인 SVM(Support Vector Machines)은 특징 공간에서 비선형 경계를 찾아 분류하는 방법으로 적은 학습 데이터에서도 좋은 분류 성능을 나타낸다고 알려져 있다. 본 논문에서는 학습데이터 수에 따른 HMM(Hidden Markov Model)과 SVM의 인식 성능을 비교하고, 최적의 특징 파라미터를 선택하기 위해 SVM을 이용하여 주성분분석과 독립성분분석을 적용하여 MFCC(Mel Frequency Cepstrum Coefficient)의 특징 공간을 변화시키면서 각각의 인식 성능을 비교 검토하였다. 실험 결과 SVM은 HMM에 비해 적은 학습데이터에서도 높은 인식 성능을 보여주었고, 독립성분분석에 의한 특징 파라미터가 특징 공간상에서의 높은 선형 분별성에 의해 다른 특징 파라미터보다 인식 성능에서 우수함을 확인 할 수 있었다.

Abstract

In this paper we propose effective speech recognizer through two recognition experiments. In general, SVM is classification method which classify two class set by finding voluntary nonlinear boundary in vector space and possesses high classification performance under few training data number. In this paper we compare recognition performance of HMM and SVM at training data number and investigate recognition performance of each feature parameter while changing feature space of MFCC using Independent Component Analysis(ICA) and Principal Component Analysis(PCA). As a result of experiment, recognition performance of SVM is better than HMM under few training data number, and feature parameter by ICA showed the highest recognition performance because of superior linear classification.

Keywords : Support Vector Machines, Independent Component Analysis, Principal Component Analysis

I. 서 론

음성신호의 효과적 특징 추출(Feature Extraction)과 인식 알고리즘의 선택은 음성인식(Speech Recognition)에서 인식성능을 좌우하는 중요한 문제이다. 현재 대부분의 음성인식 시스템은 음성의 시간적 변화를 통계적으로 모델링 할 수 있는 HMM(Hidden Markov Model)

을 주로 사용하여 좋은 결과를 보이고 있다. 그러나 HMM은 일정량 이상의 학습데이터가 있어야 하며, 불충분한 학습 데이터에 대해서는 인식 성능이 급격히 저하되는 단점이 있다. 최근 Machine Learning에 관한 연구가 활발히 진행되면서 SVM(Support Vector Machines)은 고차원의 비선형 패턴 분류에 있어서 좋은 분류 성능을 나타내며 적은 학습 데이터에 대해서도 인식 성능이 뛰어나다고 알려져 있다.^[4] Machine Learning에 관련된 알고리즘들의 분류 성능은 특징 공간 내 입력 패턴들의 분포도가 얼마나 선형적 분리가 가능하냐에 달려있다.

본 논문에서는 인식 알고리즘 비교 실험으로 학습 데

* 정회원, 동아대학교 전자공학과
(Department of Electronic Engineering, Dong-A University)

** 학생회원, LG 이노텍 구미연구소
(LG Innotek Gumi R&D Center)

접수일자: 2003년7월21일, 수정완료일: 2004년4월23일

이더 량을 변화시키면서 HMM과 SVM의 인식 성능을 비교 분석하고, 효과적 특징 파라미터 선택을 위해 SVM 인식기에 대해 입력 특징 파라미터로 MFCC(Mel Frequency Cepstrum Coefficient)외에 독립성분분석(Independent Component Analysis)과 주성분 해석(Principal Component Analysis)에 의해 MFCC의 특징 공간을 변화시킨 파라미터들을 사용하여 인식 실험을 수행하고 각각의 특징 파라미터에 대한 인식 성능과 특성을 비교 분석 하였다.

본 논문의 구성은 다음과 같다. II장에서는 음성의 특징 추출에 사용한 주성분 해석과 독립성분분석 그리고 인식 알고리즘으로 사용한 SVM에 대해 기술하고 III장에서는 실험결과 비교 및 분석에 대해 기술하였으며 IV장에서 결론을 맺는다.

II. 본 론

2.1 주성분 해석

주성분 해석기법(PCA)은 입력의 선형성과 특성 식별을 이용하여 입력 데이터에서 변동량이 큰 주축(Principal Axis)을 순서대로 찾을 수 있기 때문에 데이터의 차원을 효과적으로 축소할 수 있다. 이는 패턴 인식이나 영상 처리에서 Karhunen-Loève 변환으로 잘 알려져 있다. zero-mean 특성이 있는 n 차원 신호 x

에 대해 차원 축소(식(1))와 원신호로의 복원(식(2))은 그림 1과 같다.^[2]

$$a = [x^T q_1, x^T q_2, \dots, x^T q_l] = Q^T x \quad (1)$$

$$\hat{x} = Qa = \sum_{j=1}^l a_j q_j \quad (2)$$

여기서, $n > l$ 이며 a 는 l 차원으로 축소된 신호, \hat{x} 는 축소된 신호 a 에 의해 복원된 신호이다.

식(1)(2)에서 a 와 Q 는 x 의 공분산 행렬(Covariance Matrix) R 의 고유치와 고유벡터의 일부로, R 를 구하는 방법은 식(3)에, 고유벡터와 고유치를 구하는 방법은 식(4)에 나타내었다.

$$R = \frac{1}{n} \sum_{i=1}^n (x x^T) \quad (3)$$

$$R q_i = \lambda_i q_i, i = 1, 2, \dots, n \quad (4)$$

차원 축소된 신호 a 는 중요도 순으로 나열된 n 개의 고유벡터 Q 중에 l 개만 사용하여 얻을 수 있다

본 논문에서는 GHA(Generalized Hebbian Learning)를 이용하여 구현하였다. GHA는 자율학습특성이 있으며 학습된 가중벡터 W 는 중요도 순에 의해 내림차순으로 정렬되는 특성을 가지고 있다. GHA의 학습은 식(5)와 같다.^[2]

$$\Delta w_{ji}(n) = \eta [y_j(n) x_i(n) - y_j(n) v(n)] \quad (5)$$

$$y_j(n) = \sum_{i=0}^{l-1} w_{ji}(n) x_i(n), v(n) = \sum_{k=0}^l w_{kj}(n) y_k(n)$$

학습된 가중벡터는 입력벡터 x 의 고유벡터 특성을 가지고 있으며 식(6)에 의해서 얻어진 계수 c 는 고유치 특성을 가진다. 본 논문에서는 고유치 계수 c 를 주성분해석을 이용한 특징 파라미터로 사용하였다.

$$c_j = \sum_{i=0}^{l-1} w_{ji}(n) x_i(n), c = \{c_0, c_1, \dots, c_{l-1}\} \quad (6)$$

2.2 독립성분분석

독립성분분석(ICA)은 특정신호의 생성단계에서 선형적으로 혼합되어 있는 독립신호원들을 관측데이터에서 분리해내는 방법이다. 임의의 신호는 몇 개의 확률적으

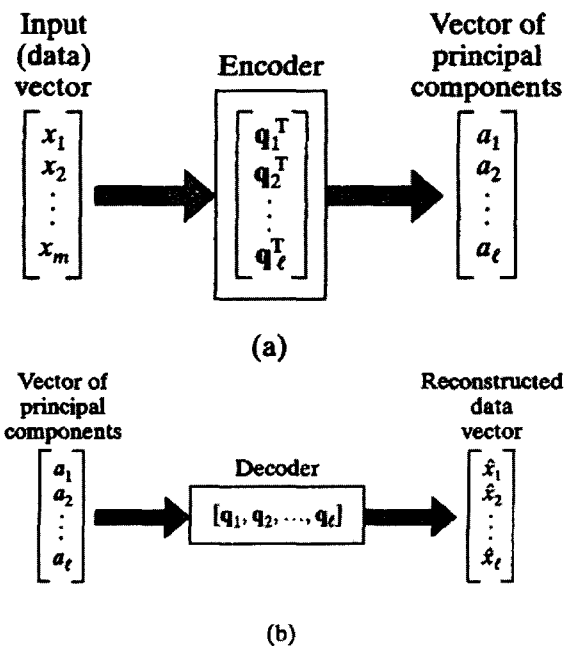


그림 1. 주성분분석 (a)차원축소 (b)복원
Fig. 1. Principal Component Analysis (a)Dimensional Reduction (b)Restoration

로 독립인 신호들에 가중치가 곱해진 다음 혼합되어 생성된 것이라 가정하고 정보이론(Information Theory), 고차 통계 특성(High Order Statistics), 추정 이론(Estimation Theory) 등을 이용하여 독립신호원간의 통계적인 의존성 최소가 되는 가중벡터를 추정하여 임의의 신호가 생성되게 하는 독립신호원을 구할 수 있다.^[1]

음성 신호는 통계적으로 독립인 고차 신호 특성들로 구성되어 있으며 고차 신호 특성들은 음성 신호의 주파수와 위상 스펙트럼을 나타내는 가중벡터를 통해서 추출될 수 있다. 음성 신호로부터 독립 고차 신호 특성들을 분리해내기 위한 가중벡터는 상호정보량(Mutual Information)을 최소화하거나 또는 비정규성(Non-gaussianity)을 최대화 하는 방법으로 학습시킬 수 있다. 이렇게 학습된 가중벡터는 음성 신호의 특정 주파수 대역에 민감한 특성을 보인다.^{[1][7]}

그림 2에서 관측벡터 \mathbf{x} 는 식(7)과 같이 서로 확률적으로 독립인 벡터 \mathbf{y} 의 선형결합형태로 표현할 수 있다.

$$\mathbf{x} = \mathbf{A}\mathbf{y} \tag{7}$$

$$\mathbf{y} = \mathbf{W}\mathbf{x} \quad \mathbf{W} = \mathbf{A}^{-1} \tag{8}$$

여기서, \mathbf{A} 는 선형혼합행렬이라 하고 독립벡터 \mathbf{y} 는 \mathbf{A} 의 역행렬인 식(8)의 가중벡터 \mathbf{W} 를 구하여 얻을 수 있다. 추정된 독립성분들의 통계적인 독립성은 상호정보량(mutual information)으로 정의한다. 관측벡터의 상호정보량은 추정된 독립성분들의 결합엔트로피와 각각의 엔트로피의 차로 계산을 하거나 Kullback-Leibler 발산정리에 의해 다음과 같이 정의한다.^{[2][5]}

$$J(\mathbf{W}) = \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod_{i=1}^N p_i(y_i)} d\mathbf{y} \\ = -H(\mathbf{y}) + \sum_{i=1}^N H(y_i) \tag{9}$$

상호정보량 $J(\mathbf{W})$ 은 항상 양수이고, 또한 각 성분들이 독립적으로 분리되었을 때는 0 이 된다.

독립성분분석의 목적은 상호정보량 $J(\mathbf{W})$ 을 최소화 하는 \mathbf{W} 를 추정하여 식(8)에 의해 독립벡터 \mathbf{y} 를 구하는 것이다. 학습방법으로 식(10)의 자연 감소법(Natural Gradient Method)을 사용하여 가중벡터 \mathbf{W} 를 추정할 수 있다.^[5]

$$\Delta \mathbf{W} = -\eta \frac{\partial J(\mathbf{W})}{\partial \mathbf{W}} [\mathbf{W}^T \mathbf{W}]$$

$$= \eta [\mathbf{I} - \boldsymbol{\varphi}(\mathbf{y})\mathbf{y}^T] \mathbf{W} \tag{10}$$

여기서 \mathbf{I} 는 단위행렬, $\boldsymbol{\varphi}(\mathbf{y}) = [\varphi_1(y_1) \cdots \varphi_N(y_N)]^T$, 그리고 $\varphi_i(y_i) = -\frac{\partial \log p(y_i)}{\partial y_i}$ 이다.

본 논문에서는 추정된 \mathbf{W} 에 의해 얻어진 독립벡터 \mathbf{y} 를 독립성분분석에 의한 특징 파라미터로 사용하였다.

2.3 Support Vector Machines

SVM은 Structural Risk Minimization(SRM)이론으로부터 발전한 비선형 이진 패턴 분류기이다.

선형 SVM에 있어서 두 개의 클래스를 구분할 수 있는 초평면(Hyperplane)은 무수히 많으나 그림 3에서 처럼 두 클래스 간 점들의 거리 ρ 를 최대화하도록 하면 OSH(Optimal Separating Hyperplane)는 유일한 해로

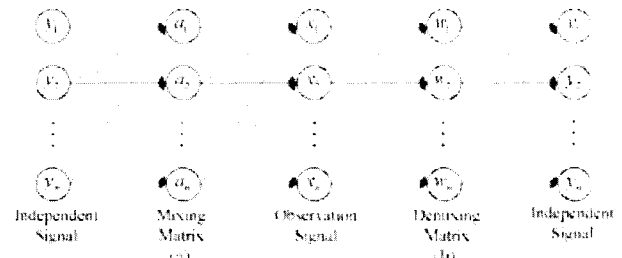


그림 2. 독립성분분석 (a) 관측신호의 생성 (b) 신호원 분리
Fig. 2. Independent Component Analysis (a) Production of Observation Signal (b) Source Signal Separation

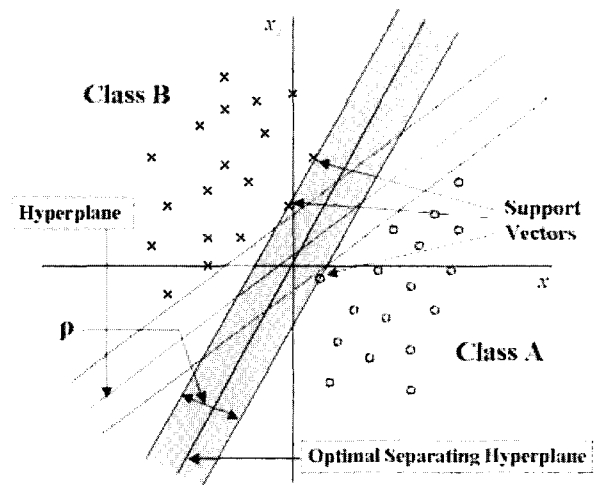


그림 3. 선형 SVM
Fig. 3. Linear SVM

존재한다.^[6]

일반적인 초평면의 방정식은 식(11)과 같고 거리 ρ (식(12))를 최대화하는 OSH의 방정식은 최적의 가중벡터 w_* 와 바이아스 b_* 로서 식(13)과 같이 표현 할 수 있다.

$$w^T \cdot x_i + b = 0 \tag{11}$$

$$\rho = 2/\|w\| \tag{12}$$

$$w_*^T \cdot x_i + b_* = 0 \tag{13}$$

거리 ρ 를 최대화 하기 위해서는 w 의 비용함수(Cost function)인 목적식(식(14))을 최소화하면서 식(15)의 제한식을 만족하는 QP(Quadratic Programming)에 의해 구한다.

$$\Phi(w) = 1/2 w^T w \tag{14}$$

$$d_i(w^T x_i + b) \geq 1 \text{ for } i = 1, 2, \dots, N \tag{15}$$

QP은 식(16)의 Lagrangian 함수의 안장점(Saddle Point)을 찾는 문제(Primal Problem)로써, 식(17)의 조건 1, 2와 식(18)의 Kuhn-Tucker 조건으로서 식(19)를 최대화 하는 α (Lagrange Multiplier)를 찾는 문제(Dual Problem)로 바꿀 수 있다.^{[2][3][4]}

$$J(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^N \alpha_i [d_i(w^T x_i + b) - 1] \tag{16}$$

$$\frac{\partial J(w, b, \alpha)}{\partial w} = 0, \frac{\partial J(w, b, \alpha)}{\partial b} = 0 \tag{17}$$

$$\alpha_i [d_i(w^T x_i + b) - 1] = 0 \tag{18}$$

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j x_i^T x_j \tag{19}$$

$$\sum_{i=1}^N \alpha_i d_i = 0, \alpha_i \geq 0 \tag{20}$$

식(20)을 만족하면서 식(19)을 최대화하는 α_* 를 구한 다음 식(21)(22)에 의해 w_*, b_* 를 구한다.

$$w_* = \sum_{i=1}^N \alpha_{*,i} d_i x_i \tag{21}$$

$$b_* = 1 - w_*^T x^{(s)} \text{ for } d^{(s)} = 1 \tag{22}$$

여기서 $x^{(s)}$ 는 그림 3의 Support Vectors이다.

결론적으로 임의의 입력 패턴 x 는 식(23)에 의해 분류된다.

$$f(x) = \text{sign}(w_*^T x + b_*) \tag{23}$$

입력 패턴이 선형 분리가 불가능한 경우는 비선형 SVM을 사용하여 Soft Margin ξ_i 과 커널(kernel) 함수 $K(x, x_i)$ 에 의해 선형분류가 가능한 고차원 공간으로 확장하여 OSH를 찾는다. 식(15)(19)에 Soft Margin 과 커널 함수를 적용하면 다음과 같다.

$$d_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \text{ for } i = 1, 2, \dots, N \tag{24}$$

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j K(x_i, x_j) \tag{25}$$

커널 함수로는 보통 다항식, RBFN(radial-basis function networks) 커널이 주로 사용되며 본 논문에서는 RBFN 커널을 사용하였고 OPC(One-Per-Class)를 사용하여 이진 분류를 다중 분류로 확장하였다.

III. 실험 및 비교 분석

3.1 인식 알고리즘 비교 실험

그림 4(a)의 실험에 사용된 음성데이터는 ETRI

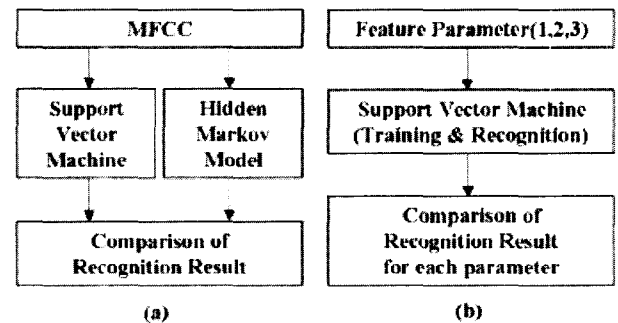


그림 4. 인식 실험 (a) 학습데이터 수에 따른 인식 알고리즘의 인식 성능 비교, (b) 특징 파라미터에 대한 인식 성능 비교

Fig. 4. Recognition Experiment (a) Recognition Performance Comparison of Recognition Algorithm for Training Data Number, (b) Recognition Performance Comparison for Feature Parameter

표 1. 분석 조건
Table 1. Analysis Condition

A/D convert	16kHz, 16bit, pre-emphasis
window	hamming(16ms, 256samples)
shifting period	3.75ms(60samples)
feature parameter	10th MFCC(except power)

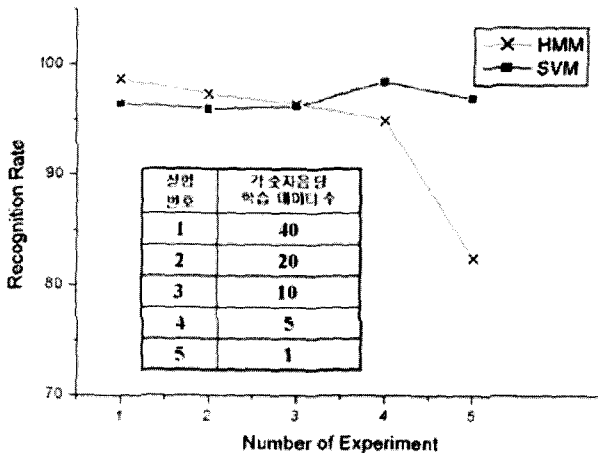


그림 5. 학습 데이터 수에 따른 인식률
Fig. 5. Recognition Rate for Training Data Number

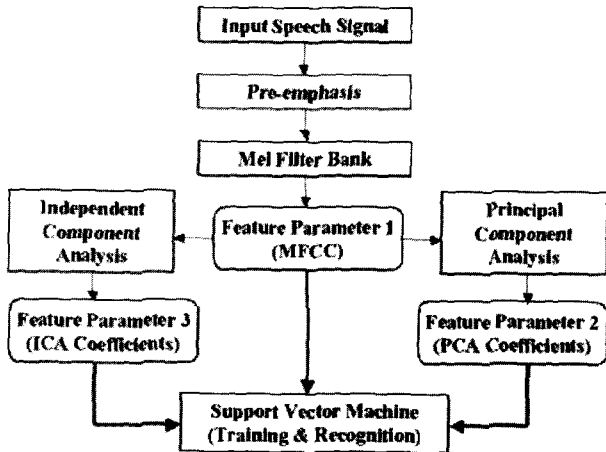


그림 6. 특징 추출 블록도
Fig. 6. Block diagram of Feature Extraction

표 2. 각 파라미터에 대한 인식률
Table 2. Recognition Rate for Each Parameter

Feature	인식 알고리즘	인식률
MFCC	SVM	96.5%
PCA Coefficient	SVM	92%
ICA Coefficient	SVM	98%

Samdori 데이터(20명의 화자가 각 숫자음 4회 발생)로 분석조건은 표 1과 같다. 인식실험은 학습 데이터 수에 따른 각각의 인식 알고리즘에 대한 인식률 변화를 관찰하였다. 결과는 그림 5와 같다. 그림 5의 결과에서 학습 데이터수가 높을 때는 HMM이 SVM보다 인식성능에서 약간의 우수함이 나타나지만 학습데이터가 적을수록 HMM에서는 알려진 바와 같이 인식률 저하가 현저히 나타나지만 SVM에서는 인식률의 변화가 거의 없었다. 위 실험의 결과로서 SVM이 적은 학습데이터에 대해서도 HMM에 비해 신용 있는 인식률을 나타냄을 알 수 있었다.

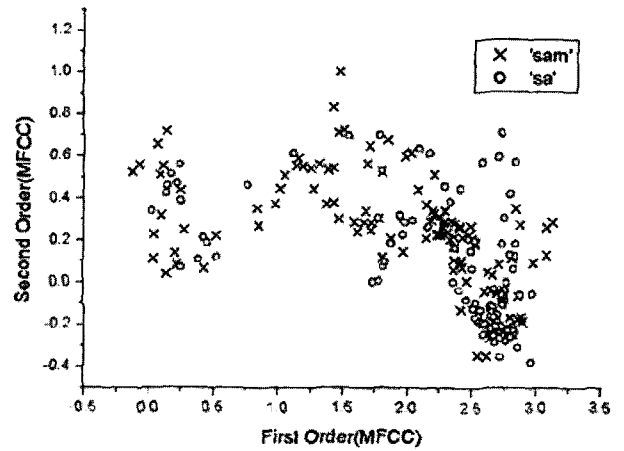


그림 7. 1,2 차 성분(MFCC)의 특징공간에서의 분포
Fig. 7. Distribution of the first & second order coefficients to the feature space(MFCC)

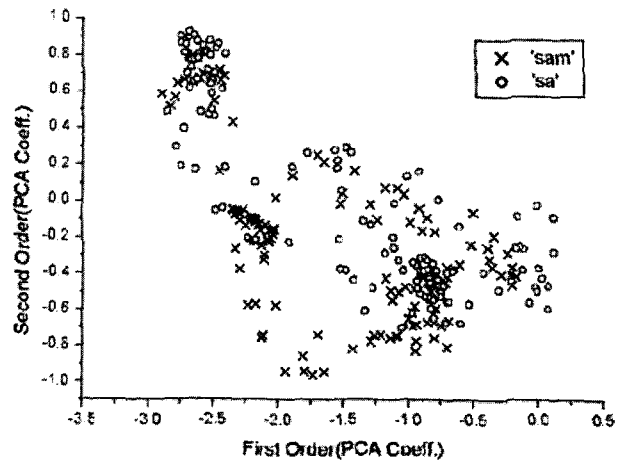


그림 8. 1, 2차 성분(PCA)의 특징공간에서의 분포
Fig. 8. Distribution of the first & second order coefficients to the feature space(PCA)

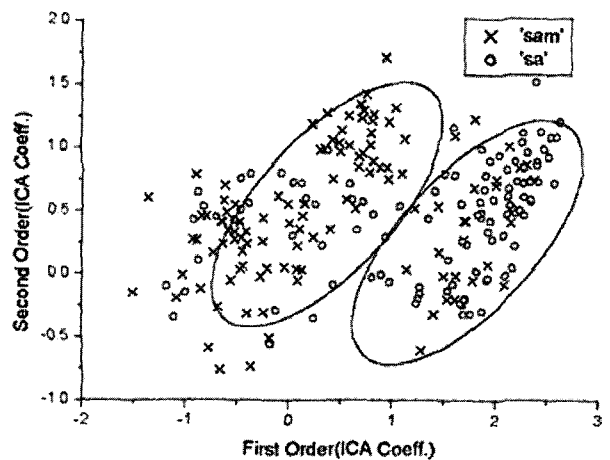


그림 9. 1, 2차 성분(ICA)의 특징공간에서의 분포
Fig. 9. Distribution of the first & second order coefficients to the feature space(ICA)

3.2. 특징 파라미터 비교 실험

그림 4(b)의 실험에 사용한 음성데이터는 그림 4(a) 실험과 동일한 ETRI Samdori 데이터를 사용하였다. 20명 화자가 2회 발성한 숫자음 400개(각 숫자음 당 40개의 학습 데이터 수)를 학습데이터로 사용하였고 인식데이터로는 학습데이터를 포함한 800개의 숫자음을 사용하였다. 각 파라미터의 특징 추출에 대한 블록도는 그림 6과 같다.

동일한 특징 차원 수에 대해 각 파라미터의 인식 성능은 독립성분분석을 통한 파라미터의 성능이 가장 우수함을 확인 할 수 있었다.

그림 7, 8, 9는 특성이 유사하여 오인식이 많은 숫자음 '삼' 과 '사'에 대한 각각의 파라미터 1차, 2차 성분만을 취하여 2차원 공간상에 분포도로 나타내었다. 그 결과 그림 9의 독립성분분석에 의한 파라미터가 특징 공간에서 선형 분별성이 가장 높다는 것을 확인 할 수 있었고 이는 SVM인식기에서 높은 성능으로 나타났다.

IV. 결 론

본 연구의 결과 제안된 음성인식 시스템은 ICA와 SVM을 사용한 시스템으로써 SVM은 학습 데이터량이 적은 상황에서도 좋은 인식성능을 보여주었고 이는 많은 학습데이터로 인한 속도 저하가 문제가 되는 응용분야에서 좋은 성능을 발휘 할 수 있을 것으로 사료된다.

그리고 독립성분분석에 의한 특징 파라미터는 MFCC나 주성분해석에 의한 특징 파라미터 보다 특징 공간상에서의 선형 분별성이 우수함을 확인 할 수 있었고 이는 다른 파라미터보다 SVM인식기에서 높은 인식 성능을 나타내었다. 차후 더 정확한 효율성 검증을 위해 화자인식, 화자 적응화 등 다양한 음성 신호 처리 분야에 적용해보아야 할 것이다.

참 고 문 헌

[1] A.Hyvärinen, J.Karhunen, E.Oja, Independent Component Analysis, John Wiley & Sons, 2001.
 [2] S.Haykin, Neural Networks, Prentice hall, 1999
 [3] Scholkopf, Smola Learning with Kernels, The MIT Press, 2002.
 [4] N.Cristianini, J.S.Taylor, An Introduction to Support Vector Machine and other kernel-based learning methods, Cambridge University Press, 2000.

[5] S.Choi, A.Cichocki, S.Amari, "Flexible independent component analysis", IEICE Trans. Fundamentals, vol.E83-A, no.12, pp.2715-2722, August, 2000.
 [6] G.J.Jang, Y.H.Oh, "Text-independent speaker identification using support vector machines", Proceeding of 11th KISS fall conference, pp.45-48, November, 1999.
 [7] C.K.Kim, S.B.Kim, S.H.Kim, K.I.Hur, "Performance Improvement of Speech Recognition Based on Independent Component Analysis", ICSP2001, Vol.2 of 2, pp.663-666, August, 2001.

저 자 소 개



김 창 근(정회원)
 1994년 2월 동아대학교
 전자공학과(공학사)
 1998년 8월 동아대학교
 전자공학과(공학석사)
 2002년 2월 동아대학교
 전자공학과 박사 수료

<주관심분야: 음성신호처리, 음성인식, DSP, Embedded System>



허 강 인(정회원)
 1980년 2월 동아대학교
 전자공학과(공학사)
 1982년 2월 동아대학교
 전자공학과(공학석사)
 1990년 2월 경희대학교
 전자공학과(공학박사)

1988년~1989년 일본 츠크바대학 정보공학부
 객원연구원
 1992년~1993년 일본 도요하시대학 정보공학부
 Post-Doc.

1984년~현재 동아대학교 공과대학 전기전자
 컴퓨터공학부 교수
 <주관심분야 : 음성신호처리, 음성인식, DSP>



박 정 원(학생회원)
 2002년 2월 동아대학교 전자공학과(공학사)
 2004년 2월 동아대학교 전자공학과(공학석사)
 2004년 1월~현재 LG 이노텍 구미연구소

<주관심분야: 음성인식, 화자인식, 수중통신, 수중 음향신호처리>