

빠른 화자 적응과 연산량 감소를 위한 MLLR알고리즘 개선

준회원 김 지 운, 정회원 정 재 호

Improvement of MLLR Algorithm for Rapid Speaker Adaptation and Reduction of Computation

Ji-Un Kim, Jae-Ho Chung

요 약

본 논문은 주성분분석(PCA, Principle Component Analysis) 혹은 독립성분분석(ICA, Independent Principle Component Analysis)를 이용하여 HMM(Hidden Markov Model) 파라메타의 차수를 감소시킴으로써 MLLR(Maximum Likelihood Linear Regression) 화자 적응 알고리즘을 개선하였다. 데이터의 특징을 잘 나타내는 PCA와 ICA를 통해 모델 mixture component의 상관관계를 줄이고 상대적으로 데이터의 분포가 적은 축을 삭제함으로써 추정해야 하는 적응 파라메타의 수를 줄였다. 기존의 MLLR 알고리즘은 SI(Speaker Independent) 모델 보다 좋은 인식성능을 나타내기 위해 30초 이상의 적응 데이터가 요구되었고, 반면 제안한 알고리즘은 적응 파라메타의 수를 감소시킴으로써 10초 이상의 적응 데이터가 요구되었다. 또한, 36차의 HMM 파라메타는 기존의 MLLR 알고리즘과 비슷한 인식성능을 나타내는 10차의 주성분이나 독립성분을 사용함으로써 MLLR 알고리즘에서 적응 파라메타를 추정할 때 요구되는 연산량을 1/167로 감소시켰다.

Keywords : Speaker adaptation, MLLR, PCA, ICA

ABSTRACT

We improved the MLLR speaker adaptation algorithm with reduction of the order of HMM parameters using PCA(Principle Component Analysis) or ICA(Independent Component Analysis). To find a smaller set of variables with less redundancy, we adapt PCA(principal component analysis) and ICA(independent component analysis) that would give as good a representation as possible, minimize the correlations between data elements, and remove the axis with less covariance or higher-order statistical independencies. Ordinary MLLR algorithm needs more than 30 seconds adaptation data to represent higher word recognition rate of SD(Speaker Dependent) models than of SI(Speaker Independent) models, whereas proposed algorithm needs just more than 10 seconds adaptation data. 10 components for ICA and PCA represent similar performance with 36 components for ordinary MLLR framework. So, compared with ordinary MLLR algorithm, the amount of total computation requested in speaker adaptation is reduced by about 1/167 in proposed MLLR algorithm.

인하대학교 전자공학과 DSP Lab.(g2001118@inhavision.inha.ac.kr, jhchung@inha.ac.kr)

논문번호: 030338-0806, 접수일자: 2003년 8월 6일

I. 서론

HMM에 기반한 음성 인식 시스템에 적용되는 화자 적응 방법은 크게 2가지로 분류된다. 그중 하나는 MAP(Maximum A Posteriori)적용 방법에 기인한 여러가지 방법이다. 이는 사전 통계로서 잘 훈련된 SI HMM 파라메타를 제공함으로써 평균, 공분산, 혼합가중치(mixture weight)등이 MAP방법에 의해 적용된다[1]. 또 다른 분류는 변환에 기초한 적응 방법으로서 HMM 파라메타를 변환 함수에 의해 변환시키는 방법이다. MLLR, SM(Stochastic Maching) 혹은 강제 변환(constrained transform)이 이 부류에 해당한다[2][9][10]. 이런 방법들은 유사한 성능을 갖는 HMM 파라메타를 군집화한 후, ML(Maximum Likelihood)을 이용하여 얻어진 변환 파라메타를 이용하여 HMM파라메타를 적응시킨다. 일반적으로 적응 데이터가 제한된다면 변환에 기반한 적응이 군집에 의존적인 변환 함수에 의해 모든 HMM 파라메타를 효과적으로 변환시킬 수 있다. 반면, 만일 충분한 데이터가 제공되면 MAP방법이 SI HMM파라메타를 변환함수에 기반한 적응방법 보다 효과적으로 적응시킬 수 있다[3][4][10].

본 논문은 적은 양의 적응 데이터가 제공되는 적응 환경을 목표로 MLLR 화자적응 알고리즘을 개선한다. MLLR 화자적응 알고리즘은 HMM의 각 혼합 성분(mixture component)들을 몇 개의 클래스로 군집화 시킨 2진 회귀나무(regression tree)를 형성하고, 적응 데이터를 각 기저 클래스에 할당한다.

만일 적응 데이터가 적게 제공될 경우, HMM 파라메타가 과도하게 적응 데이터에 편중되는 현상을 막기 위해 블록 대각 행렬 형태의 변환행렬을 이용하여 적응 파라메타 수를 감소시키거나 미리 정해진 점유 문턱값(occupation threshold)를 이용하여 2진 트리 구조의 한 회귀 클래스에서 적응의 유무를 결정한다. 그러나, 일반적으로 각 블록들은 서로 독립이라는 가정이 필요하기 때문에 음성의 특징벡터가 델타나 델타-델타를 포함한 경우에 사용이 되므로, 블록 대각 행렬의 이용이 제한적이고, 적은 적응 데이터가 제공될 경우 점유 문턱값의 설정에 따른 인식율의 변화가 매우 심한 단점이 있다. 또한, MLLR 알고리즘은 적응 파라메타를 구하기 위해 많은 양의 역행렬 연산을 수행해야 한다. 특히, 음성 특징 파라메타의 차수가 커질 수록, HMM 모델의 개수가 많아질 수록

기하 급수적으로 많은 연산량을 수행해야 한다. 예를 들어, n 차의 특징 파라메타를 사용하고 MLLR방법에서 회귀나무의 기저 클래스가 m 개라고 가정하면, $(n+1) \times (n+1)$ 차원의 행렬의 역행렬 연산을 $m \times n$ 회 행해야 한다. 따라서, 본 논문은 빠른 화자 적응과 과적응을 막기 위한 점유 문턱값의 영향의 감소를 위해 MLLR 알고리즘에서 PCA 및 ICA을 이용하여 효율적으로 HMM 파라메타의 차수를 줄임으로써 변환행렬의 적응 파라메타의 수를 감소시키고, 적응 파라메타를 추정할 때 요구되는 연산량을 감소시켰다.

본논문의 구성은 2절에서 모델 파라메타의 차수 감소를 적용한 MLLR 알고리즘을 유도하고, 3절에서는 ICA와 PCA를 이용한 모델 파라메타의 차수 감소를 설명하고, 4절에서 실험방법 및 결과를 설명한다.

II. 파라메타차수감소를 적용한MLLR

MLLR 화자적응방법은 HMM의 각 혼합 성분들을 몇 개의 클래스로 군집화 시킨 2진 회귀나무를 형성하고, 적응 데이터를 각 기저 클래스에 할당한다. 만일 적응 데이터가 적게 할당된 기저 클래스의 혼합 성분들에 대해 점유 확률이 미리 정해진 점유 문턱값보다 작으면 상위 회귀 클래스에서 다시 점유 문턱값과 비교해 혼합 성분의 적응을 결정한다. 만일, 점유 문턱값을 크게 설정할 경우, 충분한 적응 데이터가 제공되어도 하위 클래스들의 일반화된 적응 변환행렬이 추정되므로 적응 속도가 느려지고, 점유 문턱값을 작게 설정할 경우, 기저 클래스 내에서 변이가 큰 혼합 성분들의 적응 변환 행렬을 구할 때 충분한 적응 데이터가 제공되지 않으면 적응 데이터에 편중된 적응 변환 행렬이 추정되므로 인식성능이 나빠진다. 이는 적응 변환 함수를 추정할 때, 혼합 성분들의 분산이 작게 형성된 축에 대한 적응 파라메타들이 적응 데이터에 지나치게 편중되기 때문이다. 따라서, 본 논문에서는 혼합 성분들의 분산이 작게 형성된 축에 대한 적응 파라메타를 생략함으로써 추정해야하는 적응 파라메타의 수를 감소시키고 적은 양의 적응데이터에 의한 과적응의 영향을 감소시켰다. 따라서, 과적응 방지를 위한 점유 문턱값은 더욱 낮게 설정될 수 있고, 고정된 점유 문턱값에 대해 화자 적응 속도를 향상시킬 수 있다. 또한, HMM의 각 혼합 성분의 차수를 감소함으로써 연산량을 감소시킬 수 있다.

MLLR 알고리즘의 목적은 각 혼합 성분의 평균 벡터의 변환을 추정하여, SI모델을 새로운 화자의

SD(Speaker Dependent)모델로 바꾸는 것이다. 평균 μ_s 를 갖는 혼합 성분 s 에 대해 차수감소행렬과 적응행렬에 의해 변환된 평균 $\hat{\mu}_s$ 는 다음과 같이 나타낸다.

$$\hat{\mu}_s = P^{-1}W_s P \xi_s \quad (1)$$

여기서 W_s 는 차수 $(m+1) \times (m+1)$ 인 변환 행렬이고 P 는 차수 $(m+1) \times (n+1)$ 인 차수감소 행렬이다. (단, $m < n$) P^{-1} 은 차수 확장 행렬로서 차수 감소된 벡터를 원래의 차수로 변환하는 행렬이다. ξ_s 는 평균벡터를 확장시킨 것으로

$$\xi_s = \{w, \mu_{s1}, \mu_{s2}, \dots, \mu_{sm}\}' \quad (2)$$

여기서 w 는 잔류 편차 항목이다.

적응데이터를 $O = \{o_1, o_2, \dots, o_T\}$ 라고 가정하자. 여기서 T 는 관측 벡터의 개수이다. 모델 파라미터를 재추정하기 위한 보조 함수는 다음과 같이 나타낼 수 있다.

$$Q(\lambda, \bar{\lambda}) = a - \frac{1}{2} F(O|\lambda) \sum_{j=1}^T \gamma_j(o_j) [\ln \log(2\pi) + \ln |\Sigma_j| + h(o_j, j)] \quad (3)$$

$$h(o_j, j) = (o_j - P^{-1}W_j P \xi_j)' \Sigma_j^{-1} (o_j - P^{-1}W_j P \xi_j) \quad (4)$$

$Q(\lambda, \bar{\lambda})$ 를 $P^{-1}W_j$ 에 대해 일차미분을 행하면 (5)와 같다.

$$\frac{d}{dW_j} Q(\lambda, \bar{\lambda}) = F(O|\lambda) \sum_{j=1}^T \gamma_j(o_j) \Sigma_j^{-1} [o_j - P^{-1}W_j P \xi_j] \xi_j' P' \quad (5)$$

(5)의 최대값을 구하면 (6)과 같다

$$\sum_{j=1}^T \gamma_j(o_j) \Sigma_j^{-1} o_j \xi_j' P' = \sum_{j=1}^T \gamma_j(o_j) \Sigma_j^{-1} A_j P \xi_j \xi_j' P' \quad (6)$$

여기서 $A_j = P^{-1}W_j$ 이다. 식(6)의 해를 구하는 자세한 방법은 C.J. Leggetter의 논문을 참고하라[2].

기존의 MLLR 알고리즘에서는 추정해야 하는 적응 변환 행렬 W 의 크기가 $(n+1) \times (n+1)$ 임에 비하여, 제안하는 알고리즘에서는 차수 감소 행렬 P 에 의해 $(m+1) \times (m+1)$ 로 감소된다. 즉, 적응 파라미터를 추정할 때, 평균 벡터의 크기를 $(m+1) \times 1$ 로 감소함으로써 추정해야 할 적응 파라미터의 수를 $(m+1) \times (m+1)$ 로 감소시키고 각 파라미터를 추정할 때 요구되는 역행렬의 크기 또한 $(m+1) \times (m+1)$ 로 감소함으로써 요구되는 연산량을 감소한다. 따라서, 화자 적응의 성능과 연산량 감소는 차수감소행렬 P 에 의존한다. 본 논문에서는 차수감소행렬 P 를 구하기 위해 PCA 및 ICA를 이용하였다. 자세한 차수감소행렬 P 를 구하는 방법은 다음 절에서 설명한다.

III. 파라메타 차수 감소

본 논문에서는 추정해야 하는 적응 파라메타 수의 감소를 위해 주성분 분석 및 독립성분 분석을 적용한다. 주성분 분석과 독립성분 분석을 적용하기 전에 먼저 SI모델의 평균값을 몇 개의 클래스로 그룹화하는 작업이 필요한데 본 논문에서는 기존의 MLLR방법을 수행하기 전에 만들어진 M.J.F. Gales의 회귀나무[7]를 이용한다. 회귀나무를 구성하는 방법과 같고

- ①분할될 터미널 노드를 선택한다.
- ②선택된 노드에 속해 있는 혼합 성분들의 평균과 분산을 구한다.
- ③부모 노드의 평균을 기준으로 분산에 비례하여 반대 방향 두 점을 자식 노드의 평균으로 정한다.
- ④부모 노드에 속한 혼합 성분들을 Euclidean 거리를 이용하여 자식 노드에 나눈다.
- ⑤각 자식 노드에 할당된 혼합 성분들을 이용하여 자식 노드의 평균을 다시 구한다.
- ⑥평균값의 변화가 없을 때까지 ④~⑤를 반복한다.

회귀 나무를 형성한 후, 회귀 나무의 기저 클래스의 각 평균 벡터들을 연결하여 하나의 슈퍼벡터, S_i 로 만든 후 이들을 하나의 열벡터로 하는 행렬, $S = [s_1, s_2, \dots, s_k]$ 를 형성한다. 여기서 k 는 회귀 나무의 기저 클래스의 개수이다. 이와 같이 형성된 행렬 S 에 대해 주성분분석 또는 독립성분분석을 행한다.

주성분 분석의 목적은 데이터 성분들간의 상관관계를 줄이고 데이터의 특성을 잘 나타내면서 적은 차수를 갖는 변수를 찾아내는데 있다. 즉, 데이터의 분산을 최대로 하는 축을 찾아서 축 변환에 의해서 각 데이터들의 상관성을 최대로 줄임으로써 데이터를 새로운 하위영역으로 사상하는 것이다. 따라서, 각 기저 클래스의 혼합 성분들은 타 클래스의 혼합 성분과 상관관계가 감소한다. 본 논문에서 적용한 주요성분분석 방법의 순서는 다음과 같다.

- ① S^T 의 공분산 행렬 C' 를 구한다.

$$C' = E\{(S^T - E(S^T))(S^T - E(S^T))^T\} \\ = E\{(S - E(S))^T (S - E(S))\} = \frac{1}{k} \psi^T \psi \quad (7)$$

- ② $C' (= \psi^T \psi)$ 의 고유 벡터 행렬 W 와 고유값 대각 행렬 Λ 를 구한다.

$$\psi^T \psi V = \Lambda W \quad (8)$$

③양변에 ψ 를 곱한다.

$$\psi \psi^T (\psi W) = \Lambda (\psi W) \quad (9)$$

④S의 공분산 행렬C의 고유벡터 행렬과 고유값 행렬로써 ψW 와 Λ 를 사용한다.

⑤고유벡터 행렬 $V = \psi W$ 를 이용하여 혼합성분의 주성분을 구한다.

$$y = V\mu \quad (10)$$

주성분분석을 통해 얻어진 고유 벡터 행렬 $V = [v_1, v_2, \dots, v_k]$ 을 이용하여 각 혼합 성분의 평균 벡터를 고유 벡터 영역으로 사상 시킬 수 있다. 이때, 고유값에 의거하여 사용하는 고유벡터의 수를 조절할 수 있다. 작은 고유값을 갖는 고유벡터는 0에 가깝기 때문에, 작은 고유값은 생략함으로써 HMM의 평균 파라메타의 차원이 줄어들고 추정해야 하는 적응 파라메타의 수가 감소한다. 즉, 변환된 데이터의 분포의 변별력이 상대적으로 작은 고유 벡터는 사용하지 않음으로써 혼합 성분의 평균 벡터의 차수를 줄일 수가 있다. 따라서, m개의 고유 벡터를 사용한다면 변환행렬 P는 다음과 같이 나타낼 수 있다.

$$P = \begin{bmatrix} 1 & 0 \dots 0 \\ 0 & \\ \vdots & V' \\ 0 & \end{bmatrix} \quad (11)$$

여기서 $V' = [v_1, v_2, \dots, v_m]$ 이다.

독립성분분석의 목적은 주성분분석의 목적과 유사하다. 그러나, 주성분분석에서 잉여 정보량은 데이터의 각 성분들 사이의 상관관계에 의해 추정되는 반면, 독립성분분석에서는 각 성분의 독립성의 개념이 더 추가된다. 즉, 독립성분분석은 출력신호가 통계학적으로 독립인 선형 변환 시스템을 구하는 것이다. 따라서, 독립 성분 분석에 의해 각 기저 클래스 사이의 의존도와 상관 관계는 감소한다. 본 논문에서 사용한 독립성분분석 방법은 A. Hyvarinen의 FastICA로써 다음과 같다. [8]

①S의 화이트닝 변환을 수행한다.

$$Z = \Lambda^{-1/2} W^T S \quad (12)$$

②추정할 독립성분의 개수 m을 결정하고, 카운터 p를 1로 초기화한다.

③가중 행렬 A의 열벡터 a_p 의 놈(norm)을 초기화 한다.

④식 (13)을 이용하여 열벡터 a_p 를 갱신한다.

$$a_p = E\{z, g(a_p^T z_i)\} - E\{g'(a_p^T z_i)\} a_p \quad (13)$$

여기서, $g(y) = y^3$ 이다.

⑤열벡터 a_p 를 정규화한다.

$$a_p = a_p - \sum_{j=1}^{p-1} (a_p^T a_j) a_j \quad (14)$$

⑥열벡터 a_p 를 정규화한다. ($a_p \approx a_p / \|a_p\|$)

⑦열벡터 a_p 가 수렴하도록 ④~⑥을 반복한다.

⑧p를 하나 증가시키고, p가 m이 될 때까지

③~⑦을 반복한다.

주성분분석과 같은 신호의 상관관계에 기반한 변환 방법과 비교해 볼 때, 독립성분분석은 입력신호의 상관관계(2nd order statistics)뿐만 아니라 더 높은 차원의 통계학적인 의존관계를 줄인다. 따라서, HMM 파라메타의 차원을 줄일 때에는 독립성분분석이 주성분분석보다 우수한 성능을 보일 수 있다. 그러나, 입력 데이터인 HMM 파라메타에 독립성분이 충분히 포함되어 있지 않다면, 주성분분석이 독립성분 분석보다 데이터의 특징을 잘 나타낼 수 있다. [9] 주성분분석에서와 유사한 방법으로 독립성분분석에 의한 가중행렬을 이용해서 변환행렬 P를 구할 수 있다. 본 논문에서는 주성분분석에 의한 변환 행렬을 이용한 MLLR 화자적용 방법과 독립성분분석에 의한 변환 행렬을 이용한 MLLR 화자적용 방법을 기존의 선형 변환을 이용하지 않은 MLLR 화자적용 방법을 비교한다.

IV. 실험 방법 및 결과

본 논문에서 사용한 data base는 PBW(Phonetic Balanced Word) DB이다. 31명의 화자가 발성한 445개의 단어를 훈련 데이터로 사용하였다. 사용한 음성의 특징 파라메타는 18차의 LPCC(Linear Predictive Cepstral Coefficient)와 delta계수를 사용하였고, CMS(Cepstral Mean Subtraction)을 행하였다. 음성인식은 40개의 단음소(monophone)로 훈련된 모델을 이용하여 행하였다. 각 단음소 모델은 3개의 상태를 가지고 있고 각 상태는 8개의 혼합 성분을 포함한다. MLLR방법과 제안한 알고리즘의 실험에서는 블럭의 개수가 2개인 블럭 대각 변환 행렬을 사용하였다. 단, 제안한 알고리즘에서는 각각의 블럭을 따로 ICA와 PCA을 행하였다.

그림 1에는 화자독립모델의 인식성능과 기존의

MLLR, PCA나 ICA를 적용한 MLLR 알고리즘을 이용하여 화자적응 후 화자중속모델의 인식성능을 비교하였다. PCA와 ICA 모두 36차의 HMM 파라메타를 10차로 감소하였고, 화자적응의 성능에 영향을 주는 한 요소인 점유 문턱값은 고정된 값을 사용하였다. MLLR 알고리즘의 인식성능이 화자독립모델의 인식성능 보다 좋기 위해서는 30초 이상의 적응데이터가 필요한 반면, PCA나 ICA를 적용한 MLLR 알고리즘은 약 10초 이상의 데이터에 의해 적응 되었을 때 화자독립모델의 인식성능 보다 좋게 나타났다. 이는 차수 감소에 의해 필요한 적응 파라메타의 수가 감소했기 때문으로 사료된다.

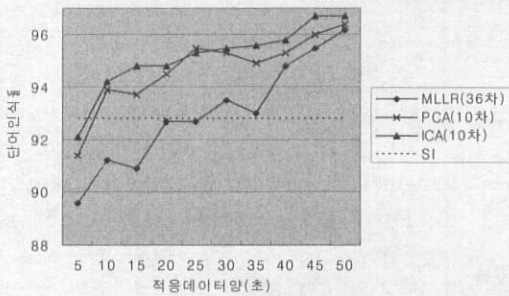


그림1. 기존의 MLLR과 PCA 혹은 ICA를 적용한 MLLR의 단어인식률 비교

그림 2에 점유 문턱값에 따른 인식 결과를 나타내었다. PCA와 ICA를 적용할 때, 독립성분과 주성분의 차수는 MLLR 알고리즘만을 적용하였을 때와 유사한 성능을 나타내는 10차를 사용하였다. 기존의 MLLR 알고리즘에서는 많은 양의 적응데이터가 제공될 경우(30초) 점유 문턱값에 따른 인식율의 최대값과 최소값의 차이는 약 1%이며, 인식율의 분산은 0.19로 점유 문턱값의 영향을 적게 받는 것으로 나타났다. 반면, 적은 양의 적응 데이터가 제공될 경우(10초) 인식율의 최대값과 최소값의 차이는 약 8.5%이고, 분산은 7.9로 나타났다. 특히, 점유 문턱값이 작아질 경우 SI 모델의 인식율 보다 낮은 인식 성능을 나타내었다.

PCA나 ICA를 적용한 MLLR 알고리즘에서는 많은 양의 적응데이터가 제공될 경우(30초), MLLR 알고리즘만을 적용하였을 때 보다 약 1% 이상 향상된 인식 성능을 나타내었고, 인식율의 분산은 ICA를 적용하였을 때 0.06, PCA를 적용하였을 때 0.11로 PCA나 ICA를 적용하지 않았을 때와 같이 점유 문턱값에 영향을 적게 받는 것으로 나타났다. 적은 양의 적응 데이터가 제공될 경우(10초), MLLR 알고리즘

만을 적용하였을 때 과적응에 의해 SI모델의 인식성능 보다 낮은 인식율을 보였던 점유 문턱값에 대해서 PCA나 ICA를 적용하였을 때 최대 6%이상 인식율이 향상되었다. 특히, ICA를 적용하였을 때 점유 문턱값에 영향을 적게 받는 것으로 나타났는데, 이는 ICA가 통계학적으로 2차원적인 분석방법인 신호의 상관관계를 줄였을 뿐 아니라, 더 높은 차원의 통계적 의존관계를 줄였기 때문으로 사료된다.

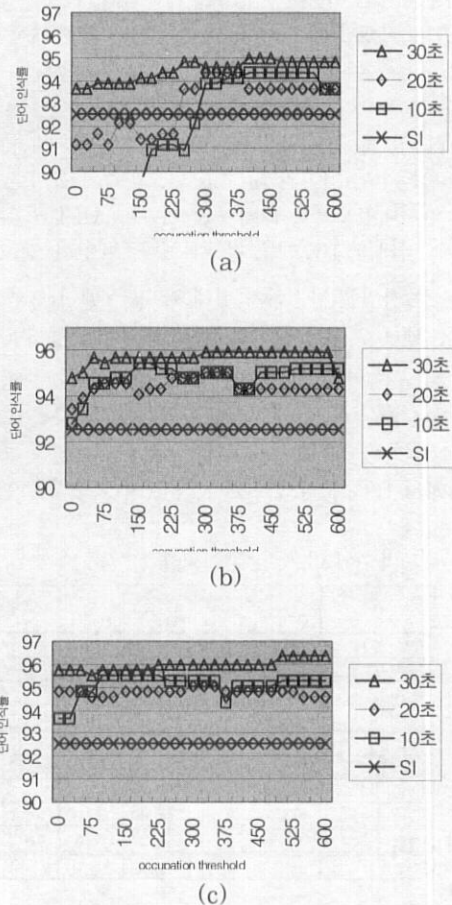


그림 2. MLLR 화자적응에서 점유 문턱값에 따른 인식율 변화. (a) 기존의 MLLR, (b) PCA를 적용한 MLLR, (c) ICA를 적용한 MLLR

표 1에 차수에 따른 단어 인식률과 기존의 MLLR 알고리즘에 대한 상대적인 연산량 감소율을 나타내었다. ICA와 PCA 모두 10차의 주성분이나 독립성분을 사용했을 때 기존의 MLLR 알고리즘의 성능과 유사한 성능을 나타내었다. 10차보다 낮은 차수에서는 적응데이터가 충분히 제공될 경우 기존의 MLLR 알고리즘보다 PCA나 ICA를 적용한 MLLR 알고리즘

의 인식성능이 낮게 나타났다. 그러나, 적응데이터가 적게 제공될 경우, PCA이나 ICA를 적용한 MLLR 알고리즘이 기존의 MLLR 알고리즘 보다 높은 인식률을 나타내었다.

연산량에 대해 고려하면, 역행렬 연산은 $O(n^3)$ 에 비례하므로 n 회의 역행렬 연산이 필요한 MLLR 화자적응을 위한 연산량은 $O(n^4)$ 에 비례한다. 본 실험에서는 하나의 변환 행렬을 구하기 위해 기존의 MLLR 알고리즘은 $(36 + 1) * (36 + 1)$ 차원의 행렬의 역행렬 연산을 36회 요구하는 반면, PCA 및 ICA를 이용한 MLLR 알고리즘은 $(10 + 1) * (10 + 1)$ 차원의 행렬의 역행렬 연산을 10 회 요구하였다. 즉, HMM 파라메타 차수를 10차로 감소하였을 때, 기존의 MLLR 알고리즘과 유사한 성능을 나타내었으므로 연산량은 기존의 MLLR 알고리즘에 비해 1/167만큼 감소하였다. 제안한 알고리즘에서는 차원을 감소하기 위해 각 평균 벡터 μ_i 와 차원 감소 행렬 P 의 행렬의 곱 연산이 추가로 요구되지만 이는 $O(n^2)$ 에 비례하므로 전체 연산량에 큰 영향을 미치지 않는다. 또한, 차원 감소 행렬을 구하기 위한 PCA와 ICA에 요구되는 연산량은 SI모형을 훈련하는 과정에 추가 되므로 화자 적응 시 요구되는 연산량은 없다.

표 1 차수에 따른 단어 인식률

	20초	40초	80초	160초	연산량 감소율
PCA(8차)	94.14	93.92	93.92	96.40	1/410
PCA(10차)	94.82	95.30	95.95	96.40	1/167
PCA(12차)	95.50	95.72	95.95	97.30	1/81
PCA(14차)	95.27	95.72	95.95	97.07	1/43
ICA(8차)	94.59	94.82	94.82	96.40	1/410
ICA(10차)	95.05	95.80	96.20	96.90	1/167
ICA(12차)	95.72	96.20	96.40	97.30	1/81
ICA(14차)	95.72	96.40	96.40	97.30	1/43
MLLR(36차)	92.50	94.37	95.27	96.62	1
SI	92.79	92.79	92.79	92.79	

V. 결론

본 논문은 PCA 혹은 ICA를 적용하여 HMM 파라메타의 차수를 감소시킴으로써 MLLR 화자 적응 알고리즘을 개선하였다. 제안한 알고리즘은 3가지 측면에서 고찰되었다. 첫째, MLLR 알고리즘에서 요구되는 적응 파라메타의 수가 감소함으로써 기존의 MLLR 알고리즘 보다 빠른 화자 적응을 나타내었다. 둘째, MLLR 알고리즘에서 적응 파라메타를 추정할 때 요구되는 역행렬 연산의 횟수가 줄었다.

PCA와 ICA에서 요구되는 연산량은 SI모형을 훈련할 때 요구되므로 화자적응시 추가되지 않고 연산량 감소 행렬을 곱하는 데에 요구되는 연산량은 역행렬 연산에 비해 상대적으로 매우 적은 양이므로 영향을 미치지 않는다. 셋째, ICA와 PCA를 적용하여 적응 데이터가 충분히 분포되지 않은 축을 삭제함으로써 과적응 방지를 위한 점유 문턱값의 영향을 감소시켰다. 적응 데이터가 충분히 제공되지 않을 경우, 기존의 MLLR 알고리즘은 점유 문턱값에 따라 인식율의 변화가 심하게 나타났고, PCA나 ICA를 적용하였을 경우 점유 문턱값에 의한 인식율의 변화는 매우 적게 나타났다.

참고 문헌

- [1] C. H. Lee, C. H. Lin, and B. H. Juang, "A study on speaker adaptation of the parameters of continuous density hidden Markov models," *IEEE Trans. on Signal Processing*, vol. 39, No. 4, April 1991, pp 806-814.
- [2] C. J. Leggetter, *Improved acoustic modelling for HMMs using linear transforms*, PhD Thesis, Univ. of Cambridge Feg. 1995.
- [3] O. Siohan, C. Chesta, and C. H. Lee, "Joint maximum a posteriori adaptation of transformation and HMM parameters," *IEEE Trans. on Speech and Audio Processing*, vol. 9, No. 4, May 2001, pp 417-428.
- [4] J. T. Chien, "Online hierarchical transformation of hidden Markov models for speech recognition," *IEEE Trans. on Speech and Audio Processing*, vol.7 No. 6, Nov. 1999, pp 656-667.
- [5] M.J.F. Gales, "The Generation and Use of Regression Class Trees For MLLR Adaptation," *TR263, Cambridge Univ.*, August, 1996
- [6] Aapo Hyvarinen, Juha Karhunen, and Erkki Oja, *Independent Component Analysis*. Willy-Interscience, 2001.
- [7] M. E. Tipping, and C. M. Bishop, "Probabilistic Principal Component

Analysis," *Journal of the Royal Statistical Society*, Series B, 61, Part 3, pp 611-612, 1999.

- [8] D. Ridder, J. Kittler, and R. P. W. Duin, "Probabilistic PCS and ICA subspace mixture models for image segmentation," *The Eleventh British Machine Vision Conference*, pp. 112-121, September, 2000
- [9] A. Sankar and C. H. Lee, "A maximum-likelihood approach to stochastic matching for robust speech recognition," *IEEE Trans. on Speech Audio Processing*, vol. 4, pp. 190-202, 1996.
- [10] V. Digalakis, "On-line adaptaion of hidden Markov models using incremental estimation algorithms," *Proc. 5th Eur. Conf. Speech Communication and Technology*, Sept. 1997, vol. 4, pp. 1859-1862.
- [11] Qiang Huo, and Bin Ma, "Online Adaptive Learning of Continuous-Density Hidden Markov Models Based on Multiple-Stream Prior Evolution and Posterior Pooling," *IEEE Trans. On Speech and Audio Processing*, Vol. 9, No. 4, May 2001, pp388-398

김 지 운(Ji-Un Kim)



1998년 2월:인하대학교 전자
공학과 졸업
2000년 2월:인하대학교 전자
공학과 석사
2000년 3월~현재:인하대학교
전자공학과 박사과정

<주관심분야> 음성인식, 화자확인, 화자적응, 화자인식.

정 재 호 (Jae-Ho Chung)



1982년 : University of
Maryland (BSEE)
1984년 : University of
Maryland (MSEE)
1990년 : Georgia Institute
of Technology (Ph.D.)

1984년 ~ 1985년 : 미국 국방성 산하 해군 연구
소, 신호처리실, 연구원

1991년 ~ 1992년 : AT&T Bell Laboratories,
음성신호처리연구실, 연구원(MTS)

1992년 ~ 현재 : 인하대학교 공과대학 전자공학
과, 정교수