한국음향학회지 제22권 제5호 pp. 360~366 (2003)

정질적 기준을 이용한 다층신경망 기반 화자증명 시스템의 등록속도 단축방법

Improving Speaker Enrolling Speed for Speaker Verification Systems Based on Multilayer Perceptrons by Using a Qualitative Background Speaker Selection

이 태 승*, 황 병 원*
(Tae-Seung Lee*, Byong-Won Hwang*)

*한국항공대학교 대학원 항공전자공학과
(접수일자: 2003년 1월 17일; 수정일자: 2003년 5월 20일; 채택일자: 2003년 5월 29일)

다층신경망 (multilayer perceptron)이 다른 패턴인식 방법에 비해 여러 가지 이점을 제공하지만 다층신경망에 기반한 화자증명 시스템은 낮은 증명오류를 달성하기 위한 대규모 배경화자로 인한 느린 등록속도의 문제를 안는다. 이 문제를 해결하기 위해 QnDCS (quantitative discriminative cohort speakers) 방법에서 화자군집 방법을 다층신경망 기반 화자증명 시스템에 도입하여 화자등록에 필요한 배경화자의 수를 줄이려는 시도가 있었다. QnDCS 방법이 목적을 어느 정도 달성하긴 했지만 등록속도의 향상률이 만족할 만한 수준이지 못했다. 본 논문에서는 보다 높은 등록속도 향상률을 달성하기 위한 방법으로서, 선택되는 배경화자의 수를 더욱 낮추는 정질에 기반한 기준을 도입한 QlDCS (qualitative discriminative cohort speakers) 방법을 제안한다. 두 방법에 대한 성능평가를 위해 다층신경망과 지속음에 기반한 화자증명 시스템과 음성 데이터베이스를 사용한 실험을 실시한다. 그 결과 제안한 방법이 QnDCS에 비해 온라인 방식의 EBP (error backpropagation)에 대한 학습속도 향상률 면에서 2배 이상 더 짧은 시간 내에 화자를 등록하는 것으로 나타나 보다 높은 효율을 지녔음을 증명한다.
핵심용어: 다층신경망, 화자증명, 화자군집, 음성인식, 패턴인식
투고분야: 음향 신호처리 분야 (1.1, 1.7), 음성처리 분야 (2.5)

Although multilayer perceptrons (MLPs) present several advantages against other pattern recognition methods, MLP-based speaker verification systems suffer from slow enrollment speed caused by many background speakers to achieve a low verification error. To solve this problem, the quantitative discriminative cohort speakers (QnDCS) method, by introducing the cohort speakers method into the systems, reduced the number of background speakers required to enroll speakers. Although the QnDCS achieved the goal to some extent, the improvement rate for the enrolling speed was still unsatisfactory. To improve the enrolling speed, this paper proposes the qualitative DCS (QlDCS) by introducing a qualitative criterion to select less background speakers. An experiment for both methods is conducted to use the speaker verification system based on MLPs and continuants, and speech database. The results of the experiment show that the proposed QlDCS method enrolls speakers in two times shorter time than the QnDCS does over the online error backpropagation (EBP) method.
Keywords: *Multilayer perceptron, Speaker verification, Cohort speakers, Speech recognition, Pattern recognition*
ASK subject classification: *Acoustic signal processing (1.1, 1.7), Speech signal processing (2.5)*

I. 서 론

화자인식에서는 사용자 편의도를 고려할 때 실시간 인식 성능뿐만 아니라 실시간 화자학습 및 적응성능도 중요하다. 보안기능이 일상 생활에서 적용되기 위해서는 엄격한 인식률과 함께 빈번한 사용에 따른 신속한 인증처리가 중요하게 고려되어야 한다. 다층신경망의 신속한 인식동작은 이와 같은 빠른 인증처리 요구를 만족시킬 수 있다. 그러나 화자인식 시스템의 사용자 편의 기준에는 신속한

책임저자: 이태승 (thestaff@hitel.net)
412-791 경기도 고양시 덕양구 화전동 200-1
한국항공대학교 대학원 항공전자공학과
(전화: 019-208-4667; 팩스: 02-3159-9986)

인증처리뿐 아니라 신속한 화자등록도 포함된다. 대부분의 사용자가 보안서비스의 등록 후 곧바로 서비스 받기를 원할 것인데, 이 때 긴 시간 동안 대기해야 한다면 서비스 등록을 포기할 가능성이 있기 때문이다. 한편, 같은 화자의 목소리라 하더라도 시간이 흐름에 따라 노화나 질병으로 인한 변화가 일어난다. 이 문제를 극복하기 위해 많은 화자인식 알고리즘에서 최근 화자발성을 이용한 적응기능을 도입하고 있다[1-3]. 적응은 학습의 확장으로 볼 수 있으므로 이 경우에도 신속한 학습은 중요하다.

파라메트릭 기반 시스템과 달리 다층신경망 기반 시스템에서는 신원증명에 필요한 계산을 신속하게 수행하는데 비해 화자등록은 많은 시간을 요구한다[4,5]. 다층신경망은 하나의 입력계층, 0개 이상의 은닉계층, 하나의 출력계층으로 구성된다. 입력계층에서는 패턴을 입력받고 은닉계층에서는 신경망의 행동능력을 결정하며 출력계층에서는 패턴입력에 대해 처리된 결과를 출력한다. 패턴인식의 경우 이 출력치는 모델의 유사도와 대응한다. 각 계층은 하나 이상의 계산노드로 구성되며 각 계층의 모든 노드는 인접한 계층의 노드와 완전연결된다. 이와 같은 구조로 인해 출력노드는 입력과 은닉계층의 노드를 공유하여 낮은 계산능력을 지닌 장치에서도 빠른 증명처리를 보여준다. 그러나 다층신경망 학습에서 모든 출력노드의 학습 목표치를 달성하기 위해 노드 사이의 가중된 내부 연결의 최적치를 결정하는 일은 쉽지 않은 문제이다. 더구나 다층신경망에서 등록화자를 학습하기 위해 요구되는 배경화자의 수는 이 문제를 더욱 가중시킨다. 이 두 문제는 화자의 등록속도를 늦추는 주요한 원인이다.

이 두 문제 가운데 다층신경망의 대규모 학습데이터 문제를 해결하기 위해 화자군집 개념을 다층신경망 기반 화자증명 시스템에 도입하여 화자등록에 필요한 배경화자의 수를 줄이는 DCS (discriminative cohort speakers) 방법이 제안되었다[6]. DCS 방법에서는 파라메트릭 방식으로 화자증명에서 제안된 화자군집 방법을 MLP 기반의 화자증명 시스템에 도입하여 화자의 등록에 필요한 배경화자의 수를 감소시키고자 하였다. 그러나 이 방법이 위에서 언급한 다층신경망의 이점을 잃지 않으면서 배경화자의 수를 줄이려는 목표를 어느 정도 달성하긴 했으나, 속도증가의 수준은 DCS의 우수함을 주장하기에는 미약한 편이었다.

본 논문에서는 DCS를 더욱 개선한 방법을 제안한다. DCS에서는 등록화자와의 비교학습을 위해 선택되는 배경화자의 수를 사전에 결정한다. 그러나 배경화자 분포의 밀도는 모든 곳에서 균일하지 못하며 등록화자의 위치에 따라 인접하는 배경화자의 수가 변동한다는 점 때문에 이와 같은 화자수 고정방식은 비효율적이다. DCS의 이러한 화자수 결정기준을 정량적 기준으로 생각할 수 있다. 이에 비해 본 논문에서 제안하는 방법은 유사성에 기반을 두고 배경화자를 선택한다. 따라서 등록화자가 높은 밀도의 배경화자 분포에 위치하는 경우에는 많은 배경화자가 선택되고 그렇지 않은 곳에 위치하면 적은 배경화자가 선택되도록 할 수 있다. 결과적으로 이와 같은 정질적 기준은 전반적으로 보다 적은 배경화자를 다층신경망 학습에 사용하여 화자가 더 짧은 시간 안에 다층신경망 기반 화자증명 시스템에 등록하게 할 수 있다.

II. 정량적 DCS

다층신경망 기반 화자증명에서 배경화자를 줄일 수 있는 가능성은 비교학습의 인접성에서 비롯된다. 임의의 등록화자가 배경화자 사이에 위치한다면 등록화자를 학습하기 위한 다층신경망의 결정 경계선은 등록화자와 인접한 배경화자에 대해서만 영향을 받는다. 이를 그림 1에서 설명한다. 이 그림에 의하면 등록화자와 인접하지 않고 멀리 떨어진 배경화자는 등록화자의 학습에 실질적으로 기여하는 바가 없다. 실제 음성 데이터는 그림에서 보는 2차원보다는 훨씬 큰 차원을 갖고 있으므로 완전히 인접하지 않는 배경화자의 비율은 그리 높지 않을 것이다. 그러나 높은 인식률 달성을 위해 상당히 많은 배경화자가 준비되어 있다면 인접하지 않은 배경화자의 비율이 상승함으로써 다층신경망의 결정 경계선 학습에 필요한 배경화자를 줄일 수 있다.

파라메트릭 기반 화자증명의 화자군집 방법에서 등록화자와 유사한 배경화자를 선택하는 작업은 아래의 방법으로 실현된다.

$$S_{Cohort} = Sel_{1,N,I}(Sort_{Dec}(P(X \mid S_{BG}))),$$
$$S_{BG} = \{S_i \mid 1 \le i \le I\} \quad (1)$$

여기서, $\mathbf{X}$는 등록화자의 음성신호이고, S_{BG}는 전체가 I명인 배경화자 집합을 나타내며, $Sort_{Dec}$는 주어진 값 집합의 요소를 내림차순으로 정렬하는 함수이고, $Sel_{1,N,I}$는 주어진 I개 요소의 집합에서 0부터 N번째까지의 요소의 배경화자를 선택하는 함수를 가리키며, S_{Cohort}는 선택된 배경화자의 집합을 나타낸다. 이에 비해 다층신경망 기반의 화자증명에서는 같은 작업을 아래와 같이 실현

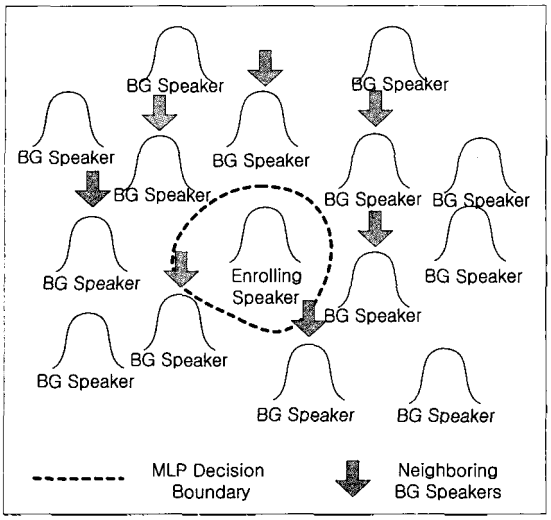

그림 1. 다층신경망의 결정 경계선 학습과 배경화자와의 관계
Fig. 1. Relationship of MLP learning of decision boundary to background speakers.

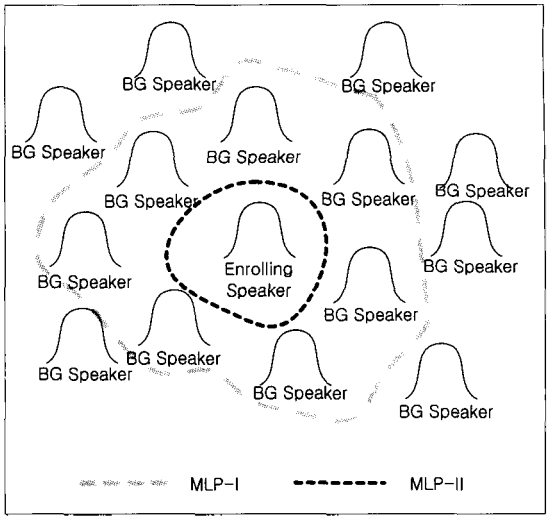

그림 2. MLP-I과 MLP-II를 이용한 QnDCS의 동작원리
Fig. 2. Function of the QnDCS using MLP-I and MLP-II.

한다.

$$S_{Cohort} = Sel_{1,N,I}(Sort_{Dec}(M_{MLP}(S_{BG} \mid \mathbf{X}))),$$
$$S_{BG} = \{S_i \mid 1 \le i \le I\} \quad (2)$$

여기서, M_{MLP}는 주어진 X에 대해 각 배경화자의 유사도를 평가하는 함수로서의 다층신경망을 나타낸다.

본 논문에서는 M_{MLP}를 계산하는 다층신경망을 MLP-I이라 하고, MLP-I을 거쳐 선택된 배경화자를 이용하여 등록화자를 학습하는 다층신경망을 MLP-II라 한다. 그리고 이렇듯 사전에 선택할 배경화자의 수를 결정하는 다층신경망 기반의 배경화자 선택 화자증명 방법을 QnDCS (quantitative discriminative cohort speakers) 방법으로 정의한다. 그림 2에서 DCS에서 배경화자가 선택된 뒤 이들이 등록화자의 학습에 사용되는 원리를 보인다. 여기서 주의할 점은 MLP-I는 등록화자와 배경화자만 식별하므로 출력노드가 1개이지만 MLP-II는 모든 배경화자를 각각 식별해야 하므로 출력노드가 I개라는 사실이다. 또한, 파라메트릭 기반 증명방식과 달리 MLP-II는 입력패턴을 출력모델에 대한 유사도로 전달하는 함수를 학습하고 이는 다층신경망의 내부 파라미터로 형성되는 결정 경계면의 최종적인 결정을 통해 이루어진다. 따라서 증명을 처리할 때 등록시 필요했던 배경화자가 증명계산에 다시 개입하지 않으므로 다층신경망의 고속증명 성능은 그대로 보존된다.

III. 정질적 DCS

배경화자의 규모와 선택기준에 따라 달라지겠지만 일반적으로 배경화자의 분포는 균등하지 못하다. 평범한 목소리의 화자가 등록한다면 인접하는 배경화자의 수가 많겠지만 그렇지 못한 화자라면 적은 수의 배경화자만이 유사성을 갖게 될 것이다. 따라서 최소의 증명오류를 나타내는 최소의 배경화자 선택한다 하더라도 그 수가 고정되어 있다면 이와 같은 화자분포의 불균등성을 수용하지 못하므로 비효율적일 수 있다. 즉, 평범한 목소리의 화자의 경우에는 충분한 배경화자를 선택하지만 평범하지 못한 화자의 경우에는 필요 이상으로 많은 배경화자를 선택하게 될 것이다.

본 논문에서는 식 (2)에서 배경화자에 대한 유사도를 측정하는 M_{MLP} 값을 활용하여 일정한 수준 이상의 유사도를 나타내는 배경화자만 이용하여 MLP-II를 학습하는 방법을 제안한다. 이 방법에서 설정되는 유사도는 등록화자를 배경화자가 충분히 둘러쌀 수 있는 수준이어야 하며, 이는 실험용 음성 데이터베이스를 이용하여 결정한다. 이 유사도가 높을수록 선택되는 배경화자가 적어지고 이에 따라 증명 오류율이 증가하므로, 음성 데이터베이스에 대해 최소의 오류율을 나타내는 최대의 유사도를 선택한다. 이와 같은 개념을 반영하여 식 (2)를 재구성하면 다음과 같다.

$$S_{Cohort} = Sel_{M_{MLP} \ge \theta, I}(Sort_{Dec}(M_{MLP}(S_{BG} \mid \mathbf{X}))),$$

$$S_{BG} = \{S_i \mid 1 \le i \le I\} \quad (3)$$

여기서 $Sel_{M_{MLP},\theta,I}$는 I명의 배경화자 중에서 M_{MLP} 이 사전 설정된 유사도 문턱값 θ를 넘는 배경화자를 선택하는 함수이다. 본 논문에서는 배경화자 선택에 유사도 수준을 고려한 이 방법을 QIDCS (qualitative discriminative cohort speakers) 방법으로 정의한다.

사전 설정되는 유사도 θ가 적절하기만 하다면 QIDCS는 현재 등록화자의 학습에 불필요한 배경화자의 수를 효과적으로 줄일 수 있다. 이는 보편적이지 못한 목소리를 갖는 화자의 경우 학습에 필요한 배경화자의 수가 크게 줄어듦으로써 가능해진다.

IV. 실 험

본 논문에서는 [6]에서 사용된 다층신경망에 기반한 화자증명 시스템과 음성 데이터베이스를 통해 QnDCS와 QIDCS의 성능차이를 시험한다. 이 시스템의 특징은 지속음을 기반으로 하기 때문에 어구독립, 어구종속, 어구요구의 모든 방식으로 시스템을 쉽게 적용시킬 수 있다는 것이다[7]. 본 논문에서 구현한 시스템은 구현의 편의를 위해 어구종속 방식을 채택한다.

4.1. 구현 화자증명 시스템

본 연구에서 구현한 화자증명 시스템은 입력음성에서 고립단어를 추출하고, 이 고립단어에서 한국어 지속음 (/a/, /e/, /ə/, /o/, /u/, /ㅟ/, /i/, /l/, 비음)을 인식한 다음, 각 지속음별로 MLP-I과 MLP-II를 이용하여 화자를 학습하고 증명점수를 계산한다. 이 시스템에서 이뤄지는 처리의 설명은 아래와 같다.

(1) 음성분석 및 특징추출

- 16 bit 16 kHz로 샘플링된 등록화자의 입력음성을 20 ms 오버랩시킨 30 ms 길이의 프레임으로 나눈다.
- 각 프레임에 대해 16차 Mel 간격 필터뱅크 (filter bank)[8]를 추출하여 고립단어 및 지속음 검출에 사용한다. 필터뱅크 계수는 전체 스펙트럼 포락에 미치는 성량의 영향을 제거하기 위해 1 kHz까지의 계수를 평균하여 모든 계수에서 차감한 뒤, 다시 모든 계수의 평균이 0이 되도록 조정한다.
- 각 프레임에 대해 50차의 0~3 kHz 대역 균등간격 Mel 필터뱅크를 추출하여 화자증명에 사용한다. 이 음성특징은 2차 포만트 (formant)에 더 많은 화자정보가 집중된다는 연구결과[9]에 의한 것이다. 필터뱅크 계수는 전체 스펙트럼 포락에 미치는 성량의 영향을 제거하기 위해 1 kHz까지의 계수를 평균하여 모든 계수에서 차감한 뒤, 다시 모든 계수의 평균이 0이 되도록 조정한다.

(2) 고립단어 검출 및 지속음 검출

- 각 지속음과 묵음을 화자독립 방식으로 검출하도록 학습된 MLP를 사용하여 고립단어와 고립단어 내의 지속음을 검출한다.

(3) 지속음별 등록화자 학습

- 지속음별로 검출된 전체 고립단어의 각 지속음을 MLP-I에 입력한 뒤, 출력뉴런의 수치를 평균하고, 이 평균치가 3절에서 설명한 θ보다 높은 배경화자를 선택한다.
- 지속음별로 선택된 배경화자 데이터를 이용하여 MLP-II에 등록화자를 학습시킨다.
- 지속음별로 MLP-II 학습에 사용되는 패턴수는 등록화자 당 10개씩이다.

(4) 지속음별 화자점수 평가

- 지속음별로 검출된 전체 고립단어의 각 지속음을 MLP-I에 입력한 뒤, 출력뉴런의 수치를 평균하고, 이 평균치가 3절에서 설명한 θ보다 높은 배경화자를 선택한다.
- 선택된 배경화자 가운데 (3)의 배경화자가 1명 이상 포함된 모든 지속음에 대해 MLP-II의 출력치를 평균한다.
- (3)의 배경화자가 1명 이상 포함된 지속음이 전무한 경우 의뢰화자를 거부한다.

(5) 등록어 및 화자점수 문턱값 비교

- (4)의 평균치와 사전 설정한 문턱값을 비교하여 최종적인 거부/수락을 결정한다.

4.2. 실험조건

실험에서 화자를 등록하는 MLP-II의 설정은 다음과 같다.

- 온라인 방식으로 학습한다.
- 입력 데이터는 -1.0~+1.0으로 평준화된다.

- 각 출력노드의 목표치로는 보다 신속한 학습을 위해 등록화자에 +0.9, 배경화자에 −0.9를 지정한다.
- 모든 내부변수는 학습 전 −0.5~+0.5의 임의수치로 초기화한다.
- 학습시 두 모델의 음성패턴은 교대로 다층신경망에 제시되는데 거의 대부분의 경우에 있어서 두 모델의 학습패턴수가 일치하지 않으므로 많은 쪽 패턴이 모두 제시될 때까지 적은 쪽 패턴을 반복해서 제시하여 1에폭을 채운다.
- 최대 학습에폭은 로컬 미니마에 빠지는 경우를 고려하여 1000회로 제한한다.
- 학습목표는 출력오류 에너지의 평균이 0.01 이하가 되는 것으로 하되, 조기 학습중지를 막기 위해 이 값의 변화율이 0.01 이하가 되어야 한다.

실험화자 40명을 한 명씩 차례로 등록화자와 실제화자로 사용하고 이를 제외한 나머지 39명을 사칭화자로 사용한다. 결과적으로 화자당 35회의 실제화자 시도와 1,560회의 사칭화자 시도를 평가하게 되고, 총명시도 화자가 40명이므로 전체적으로는 1,400회의 실제화자 시도와 54,600회의 사칭화자 시도를 평가하게 된다.

실험은 AMD 1.4 GHz급 컴퓨터에서 실시하였으며, 실험결과에서 오류율은 실제화자를 거부하는 오인 거부율과 사칭화자를 수락하는 오인 수락률이 같아지도록 검증문턱값을 조절한 동일오류율을 의미한다[7]. 학습패턴수는 1명의 화자를 등록하기 위해 학습된 총 패턴수를 나타내며, 학습시간은 이 패턴들을 학습하는데 걸린 실제시간을 가리킨다. 오류율, 학습패턴수, 학습시간은 동일한 다층신경망 학습조건에서 실시한 세 번의 실험결과를 평균한 수치를 기록한다.

4.3. 실험결과 및 분석

실험은 다층신경망의 표준적인 학습방법으로 사용되는 온라인 모드 오류역전파 알고리즘 (Online EBP: online error backpropagation)[5]과 QnDCS를 비교하는 실험과 QnDCS와 QlDCS를 비교하는 실험으로 구성된다. 첫 번째 실험에서 QnDCS의 성능향상 비율을 측정하고, 두 번째 실험에서 QnDCS에 대한 QlDCS의 성능향상 정도를 평가한다.

첫 번째 실험에서는 QnDCS에 대한 배경화자 수를 3명씩 감소시키면서 결과를 측정하였고, 그 결과를 그림 3에 정리하였다. 이 그림에서 학습속도는 화자군집 내의 화자수가 감소함에 따라 거의 선형적으로 향상되고 있음을 알 수 있다. 그러나 화자군집 내 화자수가 20명 미만인 경우에는 오류율이 증가하기 때문에 20명 이상의 화자에 대해서만 속도향상의 의미가 있다. 이를 고려했을 때 최대 등록속도 향상률은 27.1%로 기록되었다. 14명 미만의 화자에서는 오류율이 급격히 증가하는 현상이 나타나는데, 이는 배경화자 집단이 15명의 남성화자와 14명의 여성화자로 구성되었다는 점에 원인이 있을 것으로 판단된다. 즉, 14명 미만의 화자수에서는 등록화자와 동일한 성별을 갖는 배경화자의 수가 감소하게 되어 등록화자와 직접적으로 맞닿는 배경화자의 부족을 초래하기 때문이다. 마지막으로 살펴볼 부분은 배경화자의 수가 23명과

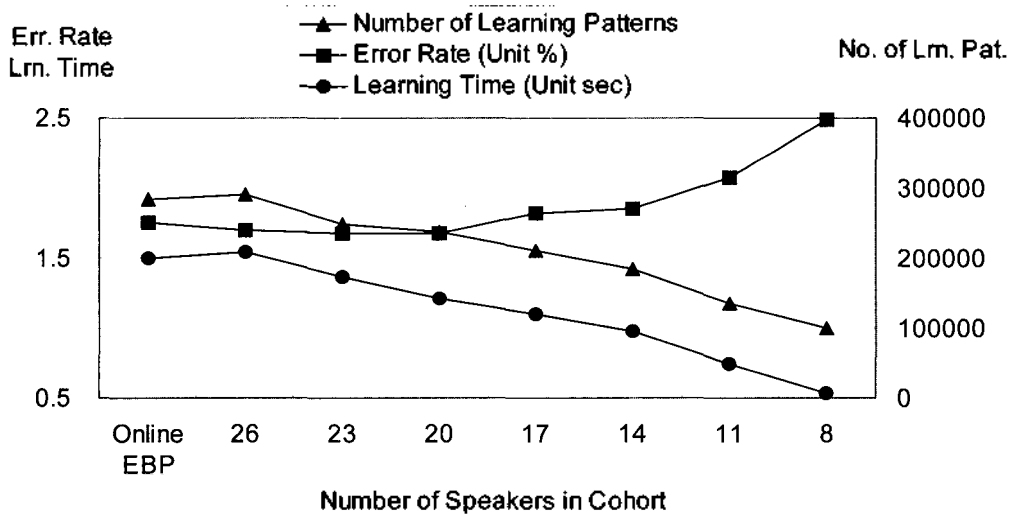

그림 3. 첫 번째 실험의 Online EBP에 대한 QnDCS의 비교결과
Fig. 3. Comparing result of the QnDCS to the online EBP in the first experiment.

표 1. 유사도 문턱값과 화자군집 내 화자수, 학습시간, 오류율 관계
Table 1. Experimental results of the QIDCS for each likelihood threshold.

Output Likelihood Threshold	Number of Speakers in Cohort	Learning Time (Sec)	Equal Error Rate (%)
-0.999	19.8	1.15	1.68
-0.995	16.1	1	1.69
-0.992	14.9	0.98	1.65
-0.99	14.3	0.93	1.76
-0.95	9.5	0.67	1.8
-0.9	7.1	0.51	2.23

20명일 때 증명 오류율이 온라인 EBP의 경우보다 낮다는 것이다. 이 결과는 화자군집 내 화자의 수가 줄어들수록 다층신경망의 교대 모델학습 특성 때문에 등록화자를 학습하는 기회가 증가하는 점에서 원인을 유추할 수 있다. 배경화자가 많을 경우 등록화자의 과도한 학습을 유발하게 되고 이것은 곧 결정 경계면을 배경화자 영역으로 부적절하게 이동시키는 것을 의미하기 때문이다.

두 번째 실험에서는 QIDCS를 적용했을 때 여러 유사도 문턱값에 대한 화자군집 내의 화자수와 오류율 및 등록화자의 학습시간을 측정한다. 배경화자 선택을 위한 MLP-의 유사도 문턱값과 이 때의 화자군집 화자수, 학습시간, 오류율을 표 1에서 정리한다. 그리고 각 화자군집 내 화자수에 대한 QIDCS의 오류율과 학습시간을 QnDCS의 경우와 비교하여 그림 4에서 나타낸다. 이 그림에서 학습시간의 감소추이는 QnDCS와 QIDCS가 거의 같은 양상을 보이지만, 오류율에 대해서는 QIDCS가 더 작은 화자수에서 비슷한 오류율을 달성하는 것을 알 수 있다. 결과적으로 Online EBP와 같은 오류율을 기록할 때의 QnDCS와 QIDCS의 화자군집 내 화자수는 각각 19명과 14.3명이며, 이 때의 학습시간은 각각 1.18초와 0.93초이다.

그림 5에서 온라인 EBP, QnDCS, QIDCS의 학습속도 차이를 정리한다. 이 결과에서 알 수 있듯이 Online EBP에 대한 학습속도 향상률이 QIDCS가 QnDCS에 비해 2배 이상 높으며, 이는 3절에서 제시한 방법이 배경화자 분포의 불균등성을 효과적으로 해결한다는 사실을 반증한다.

V. 결 론

지금까지 다층신경망 기반 화자증명 시스템의 화자등록속도를 향상시키는 방법을 고찰하였다. 다층신경망이 다른 패턴인식 방법에 비해 여러 가지 이점을 제공하지만 다층신경망에 기반한 화자증명 시스템은 낮은 증명오류를 달성하기 위한 대규모 배경화자로 인해 발생하는 느린 등록속도의 문제를 안는다. 이 문제를 해결하기 위해 QnDCS 방법에서는 화자군집 방법을 다층신경망 기반 화자증명 시스템에 도입하여 화자등록에 필요한 배경화자의 수를 줄이고자 시도하였다. 그러나 QnDCS 방법이 목적을 어느 정도 달성하긴 했지만 등록속도의 향상률이 만족할만한 수준이지 못했다. 본 논문에서는 보다 높은 등록속도 향상률을 달성하기 위한 방법으로서, 선택되는 배경화자의 수를 더욱 낮추는 정질적 기준에 기반한 QIDCS을 제안하였다. 정질적 기준의 적용에 의해 QIDCS

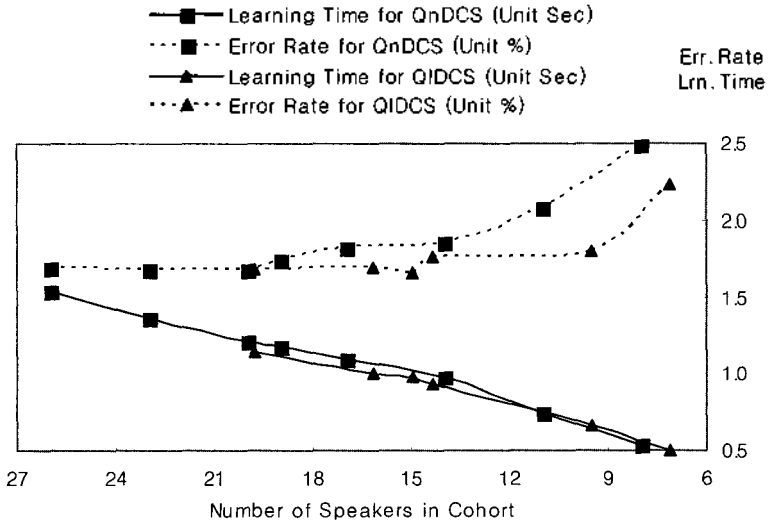

그림 4. 두 번째 실험의 QnDCS와 QIDCS 비교결과
Fig. 4. Comparison of the QIDCS with the QnDCS for the second experiment.

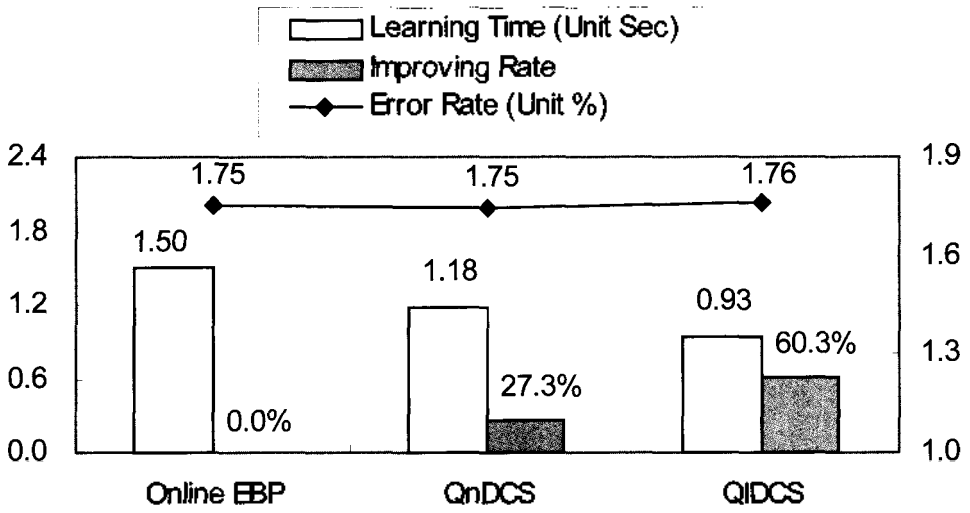

그림 5. QnDCS와 QIDCS의 학습속도 향상률 정리
Fig. 5. The learning speed improvement rates for the QnDCS and QIDCS.

는 같은 오류율 수준에서 QnDCS에 비해 더 적은 수의 배경화자를 선택할 수 있다. QIDCS의 성능을 평가하기 위해 다층신경망과 자속음에 기반한 화자증명 시스템과 음성 데이터베이스를 활용한 실험을 실시하였다. 이 실험에서 Online EBP에 대해 QIDCS가 QnDCS보다 2배 이상 높은 등록속도 향상율을 기록함으로써 QIDCS가 보다 효율적인 방법이라는 사실을 증명할 수 있었다.

참 고 문 헌

1. T. Matsui and K. Aikawa, "Robust model for speaker verification against session-dependent utterance variation," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 1, 117-120, 1998.
2. W. Mistretta and K. Farrell, "Model adaptation methods for speaker verification," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 1, 113-116 1998.
3. T. Matsui and S. Furui, "Speaker adaptation of tied-mixture-based phoneme models for text-prompted speaker recognition," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 1, 125-128, 1994.
4. A. E. Rosenberg and S. Parthasarathy, "Speaker background models for connected digit password speaker verification," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1, 81-84, 1996.
5. Y. Bengio, *Neural Networks for Speech and Sequence Recognition*, International Thomson Computer Press, 1995.
6. 이백영, 이태승, 황병원, "다층신경망 기반 화자증명 시스템에서 학습 데이터 감축을 통한 화자등록속도 향상방법," 한국음향학회지, 21 (6), 585-591, 2002.
7. S. Furui, "An overview of speaker recognition technology," *Automatic Speech and Speaker Recognition*, Kluwer Academic Publishers, 31-56, 1996.
8. C. Becchetti and L. P. Ricotti, *Speech Recognition*, John Wiley & Sons, 1999.
9. P. Cristea and Z. Valsan, "New cepstrum frequency scale for neural network speaker verification," *Proceedings of the IEEE International Conference on Electronics, Circuits and Systems*, 3, 1573-1576, 1999.

저자 약력

● 이 태 승 (Tae-Seung Lee)

1997년 2월: 한국항공대학교 항공전자공학과 학사 (공학사)
2000년 2월: 한국항공대학교 항공전자공학과 석사 (공학석사)
2000년 3월~현재: 한국항공대학교 항공전자공학과 박사과정 재학
※ 주관심분야: 음성인식, 패턴인식, 인공지능

● 황 병 원 (Byong-Won Hwang)

1972년 2월: 한국항공대학교 항공전자공학과 학사 (공학사)
1981년 2월: 도쿄대학교 전자공학과 석사 (공학석사)
1984년 2월: 도쿄대학교 전자공학과 박사 (공학박사)
1973년~1984년: 여수대학 교수
1984년~1985년: 국방과학연구소 연구원
1985년~현재: 한국항공대학교 전자정보통신컴퓨터공학부 교수
※ 주관심분야: 이미지처리, 음성인식, 패턴인식