

# 상태레벨 공유를 이용한 MLLR 적응화의 회귀클래스 생성에 관한 연구

## A Study on Regression Class Generation of MLLR Adaptation Using State Level Sharing

오 세 진\*, 성 우 창\*\*, 김 광 동\*, 노 덕 규\*, 송 민 규\*, 정 현 열\*\*  
(Se-Jin Oh\*, Woo-Chang Sung\*\*, Kwang-Dong Kim\*, Duk-Gyoo Roh\*,  
Min-Gyu Song\*, Hyun-Yeol Chung\*\*)

\* 한국천문연구원 KVN 사업본부, \*\* 영남대학교 전자정보공학부

(접수일자: 2003년 8월 19일; 수정일자: 2003년 10월 10일; 채택일자: 2003년 11월 3일)

본 논문에서는 HM-Net (Hidden Markov Network)을 다양한 태스크에의 적용과 화자의 특성을 효과적으로 나타내기 위해 HM-Net 음성인식 시스템에 MLLR (Maximum Likelihood Linear Regression) 적응방법을 도입하였으며, HM-Net 학습 알고리즘을 개량하여 회귀클래스 생성방법을 제안한다. 제안방법은 PDT-SSS (Phonetic Decision Tree-based Successive State Splitting) 알고리즘의 문맥방향 상태분할에 의한 상태레벨 공유를 이용한 방법이다. 즉, 문맥방향의 각 상태에 적응화 자 음성데이터에 포함된 문맥정보를 분할하여 적응화될 음소환경을 결정하는 것이다. 따라서 제안방법은 새로운 화자로부터 문맥정보와 적응화 데이터의 발생 양에 의존하여 결정된 많은 적응 파라미터들을 (평균, 분산) 자유롭게 제어할 수 있게 된다. 제안방법의 유효성을 확인하기 위해 국어공학센터 (KLE) 452 데이터와 항공편 예약관련 (YNU200) 연속음성을 대상으로 인식실험을 수행한 결과, 음소인식, 단어인식, 연속음성인식에 대해서, 평균 34~37%, 평균 9%, 평균 20%의 성능 향상을 각각 보였다. 또한 적응화 데이터의 양에 따른 인식성능 비교에서 제안방법을 적용한 인식 시스템이 적응 데이터의 양이 적은 경우에도 향상된 인식률을 보여 MLLR 적응방법의 특성을 만족하였다. 따라서 MLLR 적응방법을 도입한 HM-Net 음성인식 시스템에 제안한 회귀클래스 생성방법이 유효함을 확인할 수 있었다.

**핵심용어:** 상태레벨 공유, PDT-SSS 알고리즘, 회귀 클래스, MLLR 적응화

**주요분야:** 음성처리 분야 (2.5)

In this paper, we propose a generation method of regression classes for adaptation in the HM-Net (Hidden Markov Network) system. The MLLR (Maximum Likelihood Linear Regression) adaptation approach is applied to the HM-Net speech recognition system for expressing the characteristics of speaker effectively and the use of HM-Net in various tasks. For the state level sharing, the context domain state splitting of PDT-SSS (Phonetic Decision Tree-based Successive State Splitting) algorithm, which has the contextual and time domain clustering, is adopted. In each state of contextual domain, the desired phoneme classes are determined by splitting the context information (classes) including target speaker's speech data. The number of adaptation parameters, such as means and variances, is autonomously controlled by contextual domain state splitting of PDT-SSS, depending on the context information and the amount of adaptation utterances from a new speaker. The experiments are performed to verify the effectiveness of the proposed method on the KLE (The center for Korean Language Engineering) 452 data and YNU (Yeungnam Univ) 200 data. The experimental results show that the accuracies of phone, word, and sentence recognition system increased by 34~37%, 9%, and 20%, respectively. Compared with performance according to the length of adaptation utterances, the performance are also significantly improved even in short adaptation utterances. Therefore, we can argue that the proposed regression class method is well applied to HM-Net speech recognition system employing MLLR speaker adaptation.

**Keywords:** State level sharing, PDT-SSS algorithm, Regression class, MLLR adaptation

**ASK subject classification:** Speech signal processing (2,5)

## I. 서론

HMM (Hidden Markov Model)과 같은 통계적 방법에 기반한 음향모델에 있어서 모델의 구조의 결정은 매우 중요한 문제라고 할 수 있다. 음성인식에 있어서 불특정 화자 음향모델의 구조에는 다음과 같은 두 가지가 요구된다. 첫 번째로, 학습 데이터를 효율적으로 표현하는 모델의 구조에 있다. 최근 음성 데이터베이스의 확충에 따라 불특정 화자 음향모델의 작성에 대량의 학습 데이터가 사용되고 있지만 이는 통상적으로 제한된 양의 데이터라고 할 수 있다. 따라서 이것을 가능한 효과적으로 사용하고 보다 높은 성능을 얻을 수 있는 모델의 구조를 준비할 필요가 있다. 두 번째로, 소량의 데이터에도 적응화(화자, 잡음, 발성방식 등)의 수행이 쉬운 모델의 구조가 필요하다. 예를 들어, 불특정 화자 음향모델은 특정화자의 음성 데이터를 가지고 화자적응화를 수행하는 경우가 많이 있다. 고속 화자적응화를 실현하기 위해서는 소량의 음성 데이터에도 모델 전체를 목적으로 하는 화자의 특성에 적용할 수 있는 모델 구조, 즉 적응에 유리한 구조를 준비하는 것도 필요하다.

가장 이상적인 HMM을 모델의 구조 관점에서 생각해 보면 최소한의 모델 파라미터를 가지면서 성능이 우수한 모델이라고 할 수 있다. 즉, 모델의 파라미터를 필요이상으로 증가시키면 모델의 구조가 복잡하게 되어 모델의 학습에 문제가 되므로 인식성능을 저하시키는 원인이 된다. 또한 파라미터의 수가 많은 모델은 적응화할 때에도 대량의 적응용 데이터가 필요하게 된다. 따라서 위와 같은 문제를 해결할 수 있는 적절한 제약조건을 주어야 하며 이것을 모델에 대해 적절한 정도로 주어지는 것이 중요하다.

모델에 제약을 주는 한가지 방법으로 모델의 구조를 공유하는 것이다. 공유는 통계적인 성질이 유사한 여러 개의 파라미터 사이에 결합(tying) 관계를 가지도록 하는 것을 말한다. 공유의 한가지 큰 장점은 모델의 학습효율을 높일 수 있다는 것이다. 일반적으로 여러 개의 파라미터 중에서 유사한 성질을 가지고 있는 것끼리 하나로 묶어서 학습 데이터로부터 파라미터를 추정하게 된다. 공유된 파라미터의 학습 데이터는 서로 통합되고 통계적으로 보다 신뢰성이 높은 파라미터 추정을 수행할 수 있게

된다. 따라서 공유를 통해 모델의 파라미터 수를 줄이게 되므로 학습효율이 높아지며 강건한 모델을 작성할 수 있게 된다.

최근 모델의 공유구조는 파라미터 사이의 거리에 근거하여 수행되는 경우가 많이 있다. 즉, 인접하는 여러 개의 파라미터를 하나의 파라미터에 대표하게 된다. 이 관점으로부터 음소환경 클러스터링 (Phoneme Environment Clustering: PEC)[1]과 일반화된 트라이폰 (triphone) 모델[2] 등의 모델레벨의 공유화, HM-Net (Hidden Markov Network)과 같은 상태레벨의 공유화[3], 결합 미스처 HMM (또는 semi-continuous HMM)[4] 등의 분포레벨의 공유화가 제안되고 있다.

본 논문에서는 음향모델의 학습은 문헌 [3,10,13]에서 적용한 상태레벨의 파라미터 공유법인 HM-Net을 도입하였다. 기본적인 HM-Net의 작성은 SSS (Successive State Splitting) 알고리즘[3]을 이용하는데, 이 방법은 처음부터 정밀한 음향모델을 작성하지 않고 상태의 분할과 파라미터의 추정을 반복학습에 의해 모델을 점진적으로 정밀하게 학습한다. 작성한 HM-Net은 HMM의 상태 공유관계를 상태 네트워크로 표현할 수 있다. 또한 비교적 적은 양의 음성 데이터를 이용하여 적절한 공유관계(HM-Net 구조)를 구할 수 있으며 시간방향의 상태분할에 의해 음소의 문맥환경에 의존하는 상태의 길이를 설정할 수 있는 특징이 있다. 그리고 본 논문에서는 SSS 알고리즘의 장점과 음소결정트리를 결합한 PDT-SSS (Phonetic Decision Tree-based SSS) 알고리즘[9,11,13]을 도입하였는데, 이 방법은 음소결정트리에 기반한 상태 클러스터링으로부터 상태위치에 대한 독립성을 배제하고 문맥방향과 시간방향의 상태분할을 수행하여 정밀한 문맥의존 HM-Net 음향모델의 구조를 결정할 수 있다. 즉, 미리 작성해 둔 문맥의존 음향모델의 각 상태위치마다 음소질의 집합에 의해 음소결정트리를 생성하고, PDT-SSS 알고리즘에 의해 문맥의존 음향모델의 상태열을 다시 학습하는 방법이다. 하지만 PDT-SSS 알고리즘에 의한 HM-Net은 특정 태스크에서는 우수한 성능을 보이지만 태스크가 바뀔 경우 인식을 저하를 초래한다. 이를 해결하기 위한 방법으로는 모델의 학습과정에서 PBW (Phoneme Balance Word)나 다양한 문맥정보(음소환경)를 포함한 많은 양의 음성 데이터를 이용한다면 가능하

다. 하지만 음성 데이터의 수집과 모델의 학습에 많은 시간과 경비가 소요된다.

따라서 본 논문에서는 HM-Net의 단점을 보완하기 위해 적응화 방법을 HM-Net 시스템에 채용하고, HM-Net 시스템에서 상태레벨의 공유방법의 PDT-SSS 알고리즘을 MLLR 적응방법에 이용하여 회귀 클래스의 생성방법을 제안하고자 한다. 특히, PDT-SSS 알고리즘의 문맥방향 상태분할만 회귀 클래스의 문맥상의 정보를 효율적으로 표현하기 위해 사용하였다.

본 논문의 구성은 다음과 같다. II장에서는 HM-Net과 PDT-SSS 알고리즘에 대해 기술하고, III장에서 MLLR 적응화에 대해 간략히 설명한다. IV장에서 본 논문에서 제안한 상태레벨 공유를 이용한 회귀클래스 생성방법에 대해 기술한다. V장에서는 제안방법의 유효성을 확인하기 위해 음소인식, 단어인식 실험을 수행하고 그 결과에 대해 고찰한 후, 마지막으로 VI장에서 본 논문의 결론을 맺는다.

## II. HM-Net과 PDT-SSS

### 2.1. HM-Net (Hidden Markov Network)

HM-Net은 SSS 알고리즘에 의해 HMM의 각 상태를 임의의 노드로 설정하여 네트워크로 연결한 구조로 표현되며 문맥의존 HMM의 각 상태를 서로 공유하게 된다. 각 상태는 상태번호, 가능한 문맥 클래스, 선행상태와 후행상태 리스트, 자기전이 확률과 상태전이 확률, 그리고 출력확률 분포 파라미터 등의 정보를 가지고 있다. HM-Net에서는 문맥 정보가 주어질 경우 이 문맥을 만족하는 상태를 선행상태와 후행상태 리스트의 제약 조건 내에서 서로 연결하여 이 문맥에 대한 모델을 하나로 결정할 수 있다. 이 모델은 자기 루프와 이웃하는 상태로의 천이만을 허용하는 좌에서 우로 형 HMM과 동일하며 일반적인 HMM과 마찬가지로 Baum-Welch 알고리즘[14]에 의해 파라미터를 추정할 수 있다.

HM-Net 구조결정에 사용되는 SSS 알고리즘은 모든 문맥을 나타내는 1 상태의 초기모델로부터 문맥방향과 시간방향으로 상태분할 후 자동적으로 HM-Net의 구조를 결정하는 알고리즘이다. SSS 알고리즘을 그림 1에 나타내었으며, 전체적으로 간략히 설명하면 다음과 같다. 우선 유사음소단위 (Phone Likely Units)를 기본단위로 모든 모델을 연결한 네트워크 구조의 초기모델로서 각각의 모델은 하나의 상태와 그 상태를 시단에서 종단까지

결합하여 전체 학습 데이터로부터 작성한다. 상태의 분할은 경로분할을 동반하는 문맥방향과 경로분할을 동반하지 않는 시간방향이 있는데, 출력확률의 우도에 따라 한 방향으로만 수행된다. 문맥방향으로 분할할 때는 경로분할에 동반된 각각의 경로에 할당된 문맥 클래스도 동시에 분할된다. 따라서 문맥 클래스의 분할에 포함된 모든 상태 중에서 학습 데이터에 대한 누적우도 확률이 가장 큰 쪽의 상태를 분할하도록 선택된다. 시간방향으로의 상태분할에서도 누적우도 확률이 높은 쪽 상태를 분할하도록 선택된다. 이상의 상태분할을 반복하여 HM-Net의 구조가 결정된다.

### 2.2. PDT-SSS 알고리즘[9,11,13]

본 논문에서는 음향모델의 학습과 회귀클래스 트리 생성에 한국어 음성학적 지식의 음소 질의어에 의한 음소결정트리와 SSS 알고리즘의 장점을 결합한 PDT-SSS 알고리즘을 도입하였다. PDT-SSS는 SSS 알고리즘의 문맥방향 상태분할에 음소결정트리를 결합한 것으로 HM-Net에서 새로운 상태의 모델 파라미터 공유와 학습 데이터에 출현하지 않는 미지의 문맥에 대한 학습을 수행할 수 있도록 구성되어 있다. PDT-SSS 알고리즘의 주요 내용은

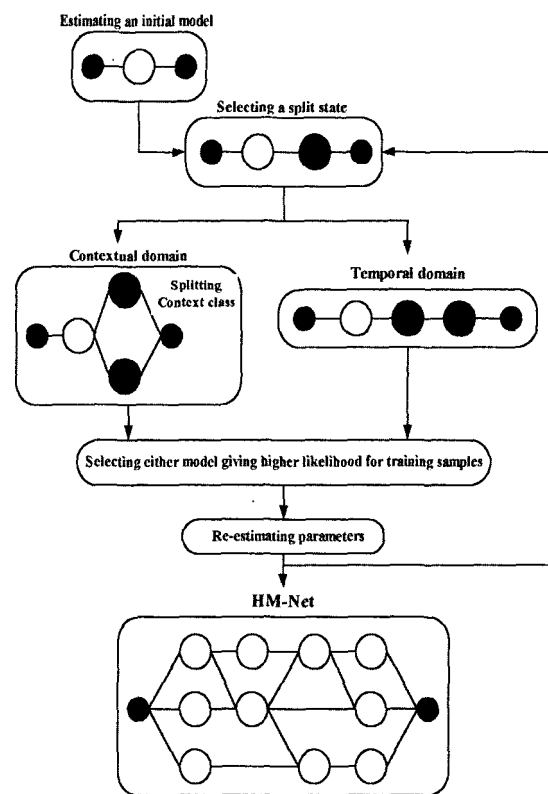


그림 1. SSS 알고리즘의 전체 구성도  
Fig. 1. Overall schematic of SSS algorithm.

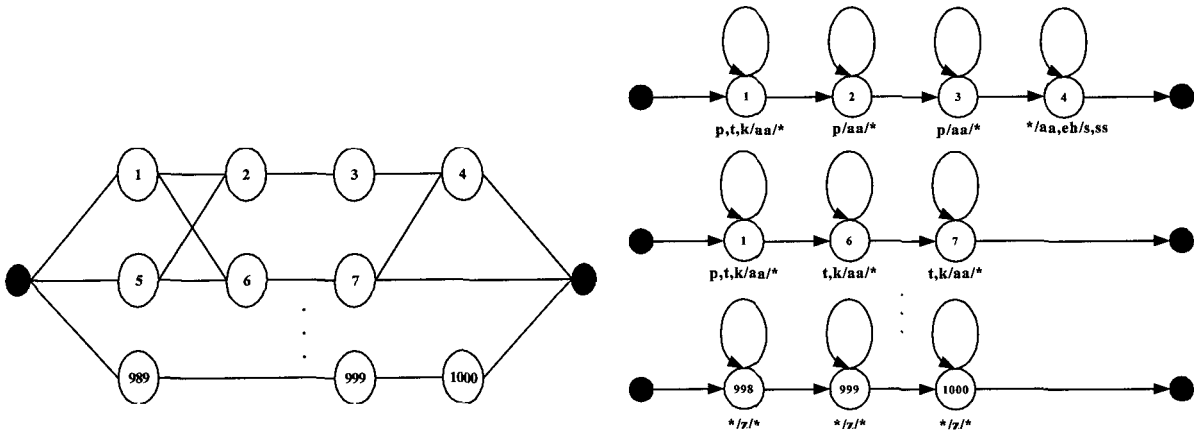


그림 2. HM-Net 모델의 예  
Fig. 2. The example of HM-Net models.

다음과 같다.

- 1) 한국어 음성학적 지식에 의한 음소 질의어 집합을 작성한다.
- 2) Baum-Welch 알고리즘으로 초기 HM-Net을 학습한다. (각 상태는 단일 가우스 분포)
- 3) 입력음성의 프레임에 따른 통계량을 계산한 후 최적 분포를 가지는 상태를 선택한다.
- 4) 문맥방향과 시간방향으로 분할할 상태를 선택한다.
  - 각 음소 질의어에 대해 문맥방향으로 분할할 때,
    - i) 질의어에 대해 허용할 수 있는 문맥 클래스의 분할과 두 개의 단일 가우스 분포를 추정한다. (각 가우스 분포는 yes 또는 no에 해당)
    - ii) 새로운 상태에 각 문맥 클래스와 각 가우스 분포를 할당한다.
  - 각 음소 질의어에 대해 시간방향으로 분할할 때,
    - i) Baum-Welch 재추정에 의해 두 개의 단일 가우스 분포를 추정한다.
    - ii) 새로운 상태에 각 가우스 분포를 할당하고 문맥 클래스를 복사한다.
- 5) 학습 샘플의 우도에 근거하여 문맥방향과 시간방향에서 최적의 HM-Net을 선택한다.
- 6) Baum-Welch 알고리즘에 의해 HM-Nets의 상태를 재학습한다.
- 7) 미리 정의한 상태수에 도달할 때까지 단계 3부터 반복한다.

그림 2에 나타난 HM-Net의 예에서 한 개의 음소를 전후해서 의존하는 트라이폰 모델의 경우 각 상태는 처리할 수 있는 문맥 클래스로서 선행/중심/후행 음소의 집합을 가지게 된다 (단, “\*”는 모든 음소의 집합을 나타냄).

그림 2에서 첫 번째 경로는 2개의 트라이폰 (p/aa/s, p/aa/ss)을 나타내는데 “p/aa/s”는 선행문맥 “p”와 후행문맥 “s”가 주어진 경우의 “aa”의 음향모델을 각각 나타낸다. 두 번째 경로는 96 개의 트라이폰 (t/aa/\*, k/aa/\*)을 나타내는데, 여기서 “\*”는 48개의 한국어 유사음소단위 중에서 임의의 하나를 의미한다[9].

### III. MLLR 적응화

MLLR 적응방법[5-7]은 원래의 모델과 적응 데이터 사이에 불일치를 줄일 수 있는 변환행렬을 이용하는 것이다. 즉, 가우시안 믹스처 HMM 시스템의 평균과 분산 파라미터들에 대한 선형 변환행렬 집합을 추정하는 것이다. 새롭게 적응된 평균을 구하는데 사용되는 변환행렬은 식 (1)과 같이 나타낼 수 있다.

$$\mu_c = A_c \zeta_c + b_c \tag{1}$$

여기서, 행렬  $A_c$ 와 바이어스 항인  $b_c$ 는 회귀 클래스  $c$ 의 MLLR 파라미터들이다. 그리고  $\zeta_c$ 는 원래 화자 독립 모델의 평균 벡터를 나타낸다. MLLR 파라미터  $A_c$ 와  $b_c$ 는 클래스  $c$ 가 최대로 되는 적응 데이터의 우도에 의해 추정된다.

회귀 클래스 트리는 회귀 클래스의 계층 또는 잎 (leaf)과 같은 기본 클래스의 집합으로 구성된다. 트리 노드 아래의 모든 기본 클래스들은 변환행렬을 공유하게 된다. 서로 다른 변환행렬들의 수는 사용할 수 있는 적응 데이터의 양에 따라 선택된다. 서로 다른 변환행렬들의 수는 이용 가능한 적응 데이터의 양에 따라서 선택되는데, 만

약 매우 적은 양의 적용 데이터가 존재한다면, 트리 루트에서의 계산된 단일 변환행렬은 모든 모델에 적용된다. 좀 더 많은 적용 데이터를 사용할 수 있다면 트리의 크기는 커지게 되며 좀 더 특정한 변환행렬을 계산할 수 있게 된다.

기본 클래스로부터 회귀클래스로의 분할은 최종 회귀 클래스 트리가 적용 데이터의 우도를 최대화 한다면 최적이라고 간주할 수 있다. 문헌[8]에서 제안한 반복 방법은 각 트리의 분할에 국부적인 우도를 최적화 함으로써 수학적으로 계산할 수 없는 최적 해를 추정하는 것이다. 이 알고리즘은 음향학적 분할로부터 얻어진 초기 회귀클래스 트리에서 시작하게 된다.

#### IV. 회귀클래스 트리의 생성방법

본 논문에서는 HM-Net의 단점을 보완하고 적응학습에 널리 사용되고 있는 MLLR 적용방법을 HM-Net에 적용하기 위해 음향모델 학습에 사용된 PDT-SSS 알고리즘의 문맥방향 상태분할 과정을 개량하여 회귀클래스 트리를 생성하도록 하였다.

우선, 음소결정트리는 음소의 음향학적 변동을 확인하기 위한 것으로 출현하지 않는 문맥의 음향학적 특성을 예측하는 방법으로 트리구조 형태의 뿌리를 문맥에 독립한 단위로 2진 트리로 표현하고 뿌리에서부터 일 방향으로 문맥 클래스의 분할을 수행한다. 이 트리는 뿌리에서 앞으로 진행함에 따라 문맥의 의존성이 강한 단위를 나타내는 계층적인 구조를 가지며 일반적으로 앞의 끝 부분에 모델을 대응시킨다. 트리의 각 접점(상태)에는 경험적으로 음소의 유사성에 기반한 질문을 하게 되며, 대답으로

yes와 no에 따라 문맥 클래스를 두 개로 분할한다. 이러한 질문은 음소 환경요인과 음소집합을 대응시킨 것도 가능하다. 임의의 문맥 클래스도 트리의 뿌리에서부터 질문을 통해 반드시 앞에 대응되어야 하기 때문에 출현하지 않는 문맥에 대해서도 음향학적으로 가장 유사한 앞에 분류될 수 있다. 그림 3에 음소 /k/의 결정트리의 예를 나타내었다.

결정트리에 대한 질의어 집합을 구성한 후 분할될 접점에 대해 최적의 상태분할을 위해 어떤 질의어가 사용될 것인가를 결정해야 한다. 결정트리의 목적은 분할될 접점의 결정에 대한 불확실성을 줄이는 것이다. 아래와 같이 회귀클래스의 생성에 사용된 PDT-SSS 알고리즘도 마찬가지로 음소결정트리의 음소 질의어에 따라 트리의 루트로부터 각 접점에 대해 반복적으로 클러스터링을 수행하여 음소결정트리를 성장시키는 방법이다. 트리의 성장은 문맥독립 HMM의 상태위치에 따라 수행된다. 기본적인 개념은 모든 분할될 접점을 트리의 뿌리에 위치시키고, 상태분할은 분할 참조에 따라 최적의 질의어 집합에 의해 반복적으로 수행된 후, 각 앞에서 결정된 상태는 서로 공유하게 된다. 본 논문에서 제안한 회귀클래스 생성을 위해 이용한 PDT-SSS 알고리즘의 장점은 상태열이 음소 질의어 집합의 yes/no의 클러스터링에 따라 학습데이터에서 출현하지 않는 음소를 가지는 모든 환경에 대해서도 구성된다는 점이다. 이와 같은 방법으로 생성된 HM-Net은 문맥독립 HMM에서 문맥방향의 상태분할에 대해서만 수행된다.

따라서 이상과 같은 점에 착안하여 본 논문에서는 PDT-SSS 알고리즘의 문맥방향 상태분할을 이용하여 MLLR 적용화 방법의 회귀클래스 생성 방법을 다음과 같이 제안한다.

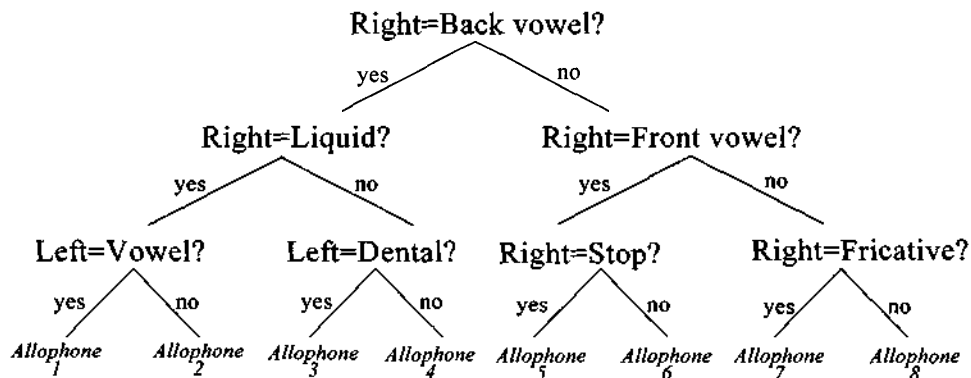


그림 3. 음소 /k/의 결정트리 예  
Fig. 3. An example of the phone /k/ decision tree.

- 1) 화자독립 학습데이터에 대해 문맥의존 음향모델 (triphone)과 상태에 대한 통계량을 학습시킨다.
- 2) 음성학적 규칙에 대한 적절한 질의어 집합을 작성하고 문맥방향의 상태분할만 있는 임의의 1 상태 HM-Net 을 작성한다.
- 3) 학습데이터에 대해 식 (2)를 이용하는데,  $d_i$  의 크기가 큰 것을 최적의 분할할 상태로 선택한다.
- 4) 각 질의어에 대해 다음과 같은 문맥방향 상태분할을 수행한다.
  - a) 질의어에 대해 두 개의 단일 가우시안 분포 (각 가우시안 분포는 yes/no와 관련)를 유도하고 분할될 문맥 클래스에 대해 상태분할을 수행한다.
  - b) 분할된 새로운 상태들에 각 문맥클래스와 각 가우시안을 지정한다.
- 5) 단계 1에서 구한 통계파일과 미리 정의한 상수 ( $\epsilon$ )와의 비교에서 상수보다 적을 경우 회귀클래스 생성을 종료한다.
- 6) 질의어의 문맥클래스에 근거한 문맥방향에서 최적의 회귀클래스를 선택한다. 여기서 단계 1의 통계량과 단계 4와 5의 단계를 통해 적응화 학습에 필요한 회귀클래스를 생성한다.

단계 3에서 분할할 상태를  $S(i)$  로 정의한다면, 분할할 상태의 선택은 입력 음성의 프레임 단위의 평균과 분산에 따라 다음 식을 사용하여 계산한다.

$$d_i = n_i \times \sum_k \frac{\sigma_{ik}^2}{\sigma_{Tk}^2} \quad (2)$$

$$\sigma_{ik}^2 = \lambda_1 \sigma_{1k}^2 + \lambda_2 \sigma_{2k}^2 + \lambda_1 \lambda_2 (\mu_{1k} - \mu_{2k})^2$$

여기서,  $K$ 는 파라미터의 차원,  $\lambda_1, \lambda_2$ 는 상태  $i$ 의 중계수,  $\mu_{1k}, \mu_{2k}$ 는 상태  $i$ 의  $k$ 번째 평균,  $\sigma_{1k}^2, \sigma_{2k}^2$ 는 상태  $i$ 의  $k$ 번째 분산,  $n_i$ 는 상태  $i$ 일 때의 학습샘플의 수, 그리고  $\sigma_{Tk}^2$ 는 모든 학습샘플의  $k$ 번째 분산을 각각 나타낸다.

단계 4에서, 이전 단계에서 선택된 상태  $S(m)$ 은  $S'(m_1)$ 과  $S''(m_2)$ 의 두 개의 상태로 분할된다. 여기서,  $m$ 은 상태번호를 나타낸다( $0 \leq m < M$ ). 우선 새로운 분포 파라미터를 구하기 위해  $S(m)$ 에 대응하는 학습 샘플을 버티비 (Viterbi) 정렬에 의해 추출하고 문맥방향의 상태분할을 수행한다. 문맥방향에서의 상태분할은 상태  $S(m)$ 을 분할하기 위한 가장 최적인 질의어의 선택과 상태  $S'(m_1)$ 과  $S''(m_2)$ 을 병렬로 연결한다. 학습 프레임 집합이 각 질의어에 의해 두 개의 집합으로 나뉘어지므로

새로운 단일 가우시안 분포는 각 집합의 평균과 분산을 계산하여 유도될 수 있다. 질의어  $q$ 가 적용됨에 따라 새로운 분포는 yes와 no에 대한  $S'(m_1)$ 과  $S''(m_2)$ 와 각각 관련이 있게 된다. 상태  $S(m)$ 에서  $C(m)$ 은 허용할 수 있는 문맥클래스로 정의하고  $c_k$ 는  $k$ 번째 요소를 나타낸다. 문맥방향 상태분할에서  $c_k$ 에 지정된 모든 프레임  $f_k$ , 평균  $\mu_k$ , 분산  $\sigma_k^2$ 을 계산하기 위해 yes와 no의 분포는  $C(m)$ 이 질의어  $q$ 에 의해 나누어질 때, 문맥의존 음향모델 (triphone)의 파라미터 공유와 음소결정트리에서의 상태분할과 같은 방법으로 다음 식들에 의해 계산된다.

$$\mu_{q, yes} = \frac{\sum_{c_k \in Q_q} f_k \mu_k}{\sum_{c_k \in Q_q} f_k} \quad (3)$$

$$\sigma_{q, yes}^2 = \frac{\sum_{c_k \in Q_q} f_k \{ \sigma_k^2 + (\mu_k - \mu_{q, yes})^2 \}}{\sum_{c_k \in Q_q} f_k} \quad (4)$$

$$\mu_{q, no} = \frac{\sum_{c_k \notin Q_q} f_k \mu_k}{\sum_{c_k \notin Q_q} f_k} \quad (5)$$

$$\sigma_{q, no}^2 = \frac{\sum_{c_k \notin Q_q} f_k \{ \sigma_k^2 + (\mu_k - \mu_{q, no})^2 \}}{\sum_{c_k \notin Q_q} f_k} \quad (6)$$

여기서,  $\mu_{q, yes}, \sigma_{q, yes}^2, \mu_{q, no}, \sigma_{q, no}^2$ 는  $C(m)$ 이 질의어  $q$ 에 대해 분할될 때 yes와 no의 평균과 분산을 각각 나타낸다. 그리고  $Q_q$ 는 질의어  $q$ 에 의해 선택되는 문맥클래스를 나타낸다.

이상과 같이 본 논문에서 제안한 회귀클래스 생성 알고리즘은 음소결정트리를 채용한 HM-Net 생성 알고리즘을 이용하고 있으므로, HM-Net 모델작성을 위한 문맥방향과 시간방향의 상태분할에 의한 학습에서, 특히 하나의 임의 음소에 대해 문맥방향 상태분할을 수행할 때 이 음소에 관련된 문맥정보 (문맥클래스)에 따라 다양한 형태의 모델 구조가 생성된다. 즉, 본래는 2진 트리 구조이지만 하나의 음소에 대해 1개 또는 여러 개의 자식 노드 (child node)를 가진 음소의 상태를 생성하게 된다. 또한, 문맥정보로 사용되는 음소 질의어에는 음성규칙이 적용되는 형태에 따라 음소 클래스로 되어 있기 때문에 하나의 임의 음소를 상태분할하면 질의어에 따라 다양한 구조의 모델이 작성된다. 따라서 이러한 점에 의해 본 논문에서 제안한 방법으로 회귀클래스를 생성하면 그림 5와 같은 구조를 가지며, 상태분할이 종료된 끝 노드는 음소 클래스를 의미한다.

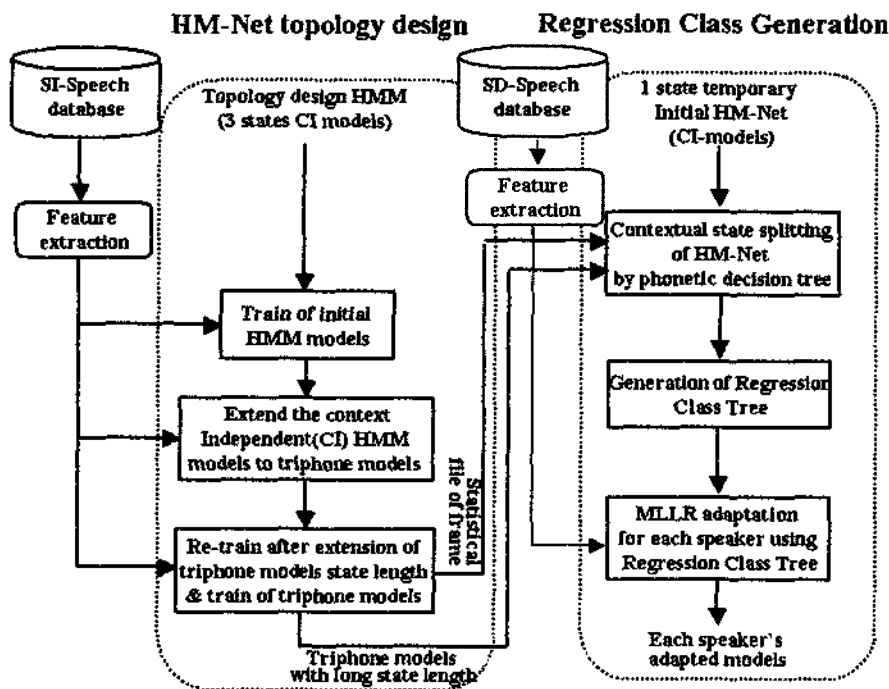


그림 4. PDT-SSS 알고리즘에 의한 회귀클래스 트리 생성의 전체 구성도  
 Fig. 4. Overall block diagram of regression class tree generation by PDT-SSS algorithm.

위에서 설명한 제안방법을 그림으로 나타내면 그림 4와 같다.

### V. 인식실험 및 고찰

본 논문에서 HM-Net의 단점을 보완하기 위해 적응화 방법을 HM-Net 시스템에 채용하고, HM-Net 시스템에서 상태레벨의 공유방법의 PDT-SSS 알고리즘에 의해 제안한 MLLR의 회귀 클래스의 생성방법의 유효성을 확인하기 위하여 음소인식, 단어인식 및 연속음성인식 실험을 수행하였다.

#### 5.1. 음성 데이터 및 분석조건

사용된 음성 데이터로는 화자독립 문맥의존 음향모델 학습에는 ETRI에서 제공된 대어휘 음성데이터 중 남녀 각 200명의 280발성을 사용하였다. ETRI에 제공된 음성 데이터는 단음절, 숫자음, 단어, 연속음성 등으로 구성되어 있어 있으며 조용한 사무실 환경에서 채록되었으며, 적응화 학습 후 평가의 경우 KLE에서 제공된 452단어 음성 데이터와 항공편 예약관련 200문장을 사용하였다.

화자독립 문맥의존 음향모델의 학습은 PDT-SSS 알고리즘의 문맥방향 상태분할을 위해 한국어 음성학적 규칙

에 의해 152개의 질의어를 작성하였다[9]. 기본 HM-Net 모델의 초기구조는 43개의 문맥 독립적인 음소모델이 병렬로 연결된 126개의 상태를 가진다. 최종 학습된 모델의 상태수는 2000개 혼합수는 4를 가지며, 이 모델을 적응화 학습의 기본모델로 사용하였다.

먼저 작성한 문맥의존 음향모델의 강건함을 평가하기 위해 태스크 독립 화자독립 단어인식 실험의 평가에 사용된 음성데이터는 KLE의 남성 35명과 여성 32명이 발성한 452단어의 1회 발성을 사용하였다.

음소 및 단어인식 실험의 경우에는 적응화 학습과 평가에 사용된 음성데이터는 KLE의 남자 35명과 여자 32명이 452단어를 2회 발성한 데이터로서 이중에서 2회 발성을 적응화 학습에 사용하고 1회 발성은 평가에 사용하였다. 연속음성인식 실험의 경우 남성 14명이 200문장을 1회 발성한 데이터 중 발성이 유사한 종류를 100문장씩 나누어 두 개의 집단으로 구성한 후, 100문장을 적응화 학습에 사용하고 나머지 100문장을 평가에 사용하였다.

적응화 학습에서 회귀클래스 트리를 위한 질의어의 경우 모델학습에서 사용된 152개의 질의어 중 76개를 이용하여 문맥의존 모델의 중심음소에 질의어가 위치하도록 구성한 후 학습하였다. 이렇게 하여 생성된 회귀클래스 트리의 모습은 음소결정트리에 의한 것과 유사하며 총 클래스의 개수는 29개로서 그림 5에 나타내었다.

모든 음성데이터는 16 kHz, 16 bits로 샘플링된 후  $1 - 0.97z^{-1}$ 의 필터를 통과시켰으며, 입력 음성의 각 프레임에 25 ms의 해밍 윈도우를 곱하여 10 ms씩 이동하면서 분석하였다. 특징 파라미터는 12차 LPC-MEL cepstrum 계수와 정규화된 대수 에너지에 1차 및 2차의 차분성분을 포함하여 총 39차를 사용하였다. 표 1과 2에 본 논문에서 사용된 음성 데이터베이스의 종류와 분석조건에 대해 각각 나타내었다.

음성인식 방법은 음소 및 단어인식의 경우 단일 경로 비터비 빔 탐색 알고리즘[14]으로서 음소인식의 경우 유사음소단위로 구성된 음소쌍 (phone-pair) 문법[14]을 사용하고, 단어인식의 경우 452단어로 구성된 단어쌍 (word-pair) 문법[14]을 사용하였다. 또한 연속음성인식의 경우 다중 경로 탐색 알고리즘[9]으로서 1-pass 탐색의 경우, 단어 2-gram 언어모델을 이용하여 프레임 동기형 비터비 빔 탐색을 수행한 후 단어 그래프를 출력한다. 2-pass 탐색의 경우 1-pass의 단어 그래프와 보다 정밀한 단어 3-gram 언어모델을 이용하여 A\* stack decoding 탐색을 수행한 후 그 결과에 대해 재스코어링에 의해 인식결과를 출력한다.

### 5.2. 적응화 학습 전 태스크 독립 실험

음성인식 시스템을 사용하는 사용자의 환경에 강건한 음향모델을 작성하기 위해 도입한 적응화 학습의 유효성을 확인하기 위해 적응화 학습 전 모델의 강건함을 평가하기 위해 태스크 독립 음성인식 실험을 수행하였다.

기본 문맥의존 음향모델은 ETRI의 대어휘 음성데이터를 이용하여 작성하였으며, 상태수 1000, 1500, 2000개와 혼합수 1, 2, 4, 6, 8개를 각각 가진다. 작성한 문맥의존 음향모델의 강건함을 평가하기 위한 태스크 독립 단어인식 실험은 기본 음향모델과 발성환경과 발성리스트가 서로 다르며, KLE의 남성 35명과 여성 32명이 발성한 452 단어를 대상으로 하였다. 기본 음향모델의 HM-Net 상태수/혼합수의 차이에 따른 태스크 독립 단어인식 실험결과를 그림 6과 7에 각각 나타내었다.

그림 6과 7의 단어인식 결과에서 모델의 상태수와 혼합수가 증가할수록 인식률의 향상은 보이지만, 모델의 작성에 사용된 음성데이터의 양에 비하여 전체적으로 비교적 낮은 인식률을 보인다. 기존의 HM-Net 음성인식 시스템에 대한 평가에서 태스크 종속 인식결과와 태스크 독립에 비해 성능이 우수함을 확인하였다[9]. 이는 HM-Net 모델을 학습할 때 많은 문맥환경과 발성리스트

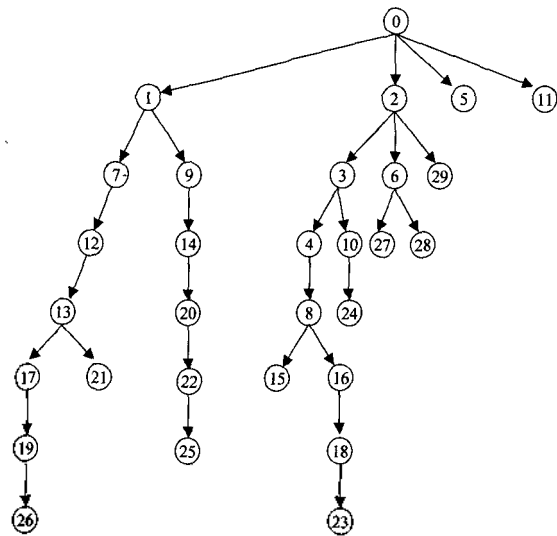
표 1. 음성 데이터베이스  
Table 1. Speech databases.

Exp.	Speaker Independent Model Construction	
Speech Data	ETRI 500	KLE 452
Train	Male/Female each 200spks. 1utt.	Male 35 spks. 1utt.
Question List	152	
Model Type	HM-Net (triphone): 1000,1500,2000states with 1,2,4,6,8 mixtures	
Exp.	Adaptation Test	
Speech Data	KLE 452	AIR 200
Adaptation Train	Male 35spks./Female 32spks. 2 utt.	Male 14 spks. each 100 sent. A
Test	Male 35spks./Female 32spks. 1 utt.	Male 14 spks. each 100 sent. B
Question List	76 (to generate regression class tree)	
Model Type	HM-Net (triphone): 2000states with 4 mixtures	
Recog. Test	Phone Recog. & Word Recog. on 452 words Sent Recog. on 100 sentences.	

표 2. 음성 데이터의 분석조건  
Table 2. Analysis conditions of speech data.

Sampling Freq.	16 kHz
Quantization	16 bit
Frame Length	25 ms
Frame Period	10 ms
Window	Hamming Window
Feature Parameters	12 order LPC-MEL cepstrum + delta power + Regression Coef. of 1, 2order = 39orders





State Transition	Question	State Transition	Question
0→1	Voiced	12→13	Arti maner fricative unasp lenis
0→2	Narrow point pharynx	14→20	Phone b1
0→5	Jaw half narrow	4→8	Tongue back rounded
0→11	Tongue mid unrounded	10→24	Arti place bilabial
1→7	Arti maner plosive unasp lenis	13→17	Arti place palatal
1→9	Phone g1	13→21	Arti maner affricate unasp fortis
2→3	Narrow point palatal	20→22	Arti place dental
2→6	Tongue high vowel	8→15	Vowel
2→29	Tongue front unrounded	8→16	Phone ju
7→12	Arti place glottal	17→19	Arti maner fricative unasp fortis
9→14	Phone d	22→25	Arti place bilabial
3→4	Nasal	16→18	Tongue back unrounded
3→10	Arti place velar	19→26	Phone p
6→27	Tongue midhigh vowel	18→23	Phone jv
6→28	Tongue front unrounded		

※ Arti: Articulation, maner: manner, unasp: Unaspirated  
 ※ Each state number represent the phoneme class.

그림 5. PDT-SSS의 문맥방향으로 음소 클러스터링에 의해 생성된 회귀클래스 트리  
 Fig. 5. The regression class tree obtained from clustering of phonemes with contextual state splitting of PDT-SSS.

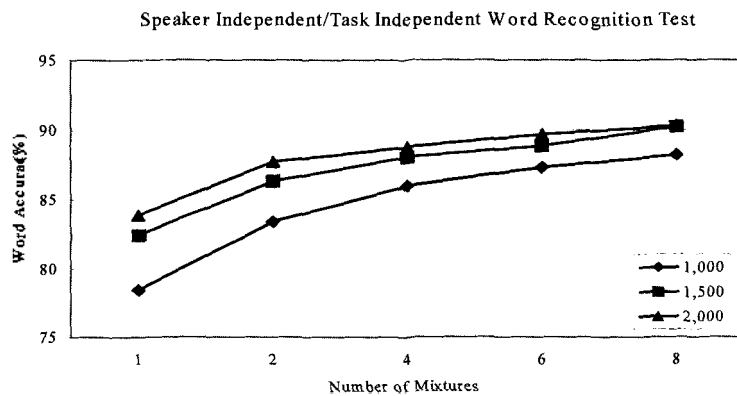


그림 6. 태스크 독립 KLE 452 남성 35명의 단어인식률  
 Fig. 6. Word accuracy of KLE452 male 35 speakers on task independent.

를 공유할 경우 높은 인식성능을 보이지만 그렇지 못한 경우에는 다소 인식성능이 저하되는 문제점으로 생각된다. 이러한 점을 보완하고 HM-Net 음성인식 시스템을 다양한 태스크에 적용하기 위해 적응화를 도입한 HM-Net 음성인식 시스템의 평가를 5.3절에 나타내었다.

**5.3. 적응화 후 음소, 단어, 연속음성인식 실험**

일반적으로 음성인식의 최종목표가 불특정 화자의 입의 발성에 대해 인식 시스템이 높은 인식성능을 가지는 것이다. 특히 최근에 소개된 HMM의 개량형인 HM-Net에 의한 인식 시스템에서 우수한 성능을 보이고 있지만 [9], 5.2절의 태스크 독립 단어인식실험의 결과와 같이 HM-Net 음성인식 시스템은 모델의 학습과 평가에 사용된 음성 데이터의 태스크가 달라지면 인식률이 저하되는 문제점을 가지고 있다. 특히 이는 HM-Net 모델을 학습할 때 많은 문맥환경과 발성리스트를 공유할 경우 높은 인식성능을 보이지만 그렇지 못한 경우에는 다소 인식성능이 저하되게 된다.

본 논문에서는 이상의 문제점 해결과 HM-Net에 의한 태스크 변화에 따른 모델 학습과정에서 효과적으로 HM-Net 음성인식 시스템을 적용하고, 제안한 MLR 적응화 방법의 화귀클래스 트리 생성방법의 유효성을 확인하기 위해 KLE 452 단어와 항공편 예약 200문장을 대상으로 음소, 단어 및 연속음성에 대해 적응화 학습을 수행한 후 인식실험을 수행하였다.

음소 및 단어인식 실험에 사용된 화자독립 음향모델은 ETRI의 대어휘 음성데이터를 이용하여 작성한 상태수 2000개 혼합수 4개인 HM-Net 문맥의존 음향모델을 사용하였으며, KLE 452 단어에서 남성과 여성의 두 번째

발성 중 적응화 단어수를 10, 50, 100, 150, 200, 452단어로 변화시키면서 학습을 수행한 후 첫 번째 발성을 평가에 사용하였다. 인식실험 결과를 그림 6과 표 3에 각각 나타내었다. 또한 항공편 예약관련 연속음성인식 실험의 경우에는 KLE의 452 단어 음성 데이터를 이용하여 작성한 상태수 2000개 혼합수 4개인 HM-Net 문맥의존 음향 모델을 사용하였으며, 각 화자의 200문장 중 100장을 적응화에 사용하고 나머지 100문장을 평가에 사용하였다. 각 100문장은 서로 유사한 발성단어를 포함하고 있으며 평가문장의 복잡도는 43.2이다.

먼저 일반적인 MLR 적응방법이 적은 적응화 데이터에서 효과적인 인식을 향상을 보임을 확인하기 위해 음성 데이터의 양을 변화시키면서 단어인식실험을 수행하였다. 그 결과, 그림 8에 나타낸 것과 같이 적응화 단어수를 조금씩 증가시킨 경우 인식성능이 증가함을 볼 수 있었다. 적응화를 수행하기 전, KLE 남성 35명과 여성 32명의 화자독립 단어인식률이 모두 88.7%를 보이지만, 적응화 학습을 200단어로 한 경우 남성 35명의 경우는 97.1%, 여성 32명의 경우는 96.9%의 평균 단어인식률을 보였다.

또한 각 화자의 전체 452단어를 적응화 학습한 경우의 음소 및 단어인식률을 표 3에 나타내었다. 음소인식의 경우 적응화 전과 비교하여 적응화 후 평균 34~37%의 인식을 향상을 보임을 확인하였다. 또한 그림 8과 표 3에서 적응단어를 200단어로 한 것과 452단어로 한 경우에서 452단어로 적응화 경우가 약 1% 향상된 인식성능을 보이지만 전체적으로 비슷한 인식성능을 보임을 알 수 있다. 따라서 MLR 적응방법의 특성인 소규모 음성 데이터에 대한 특성을 만족하는 것을 확인할 수 있었다.

그리고 표 4의 연속음성인식 결과에서도 본 논문에서

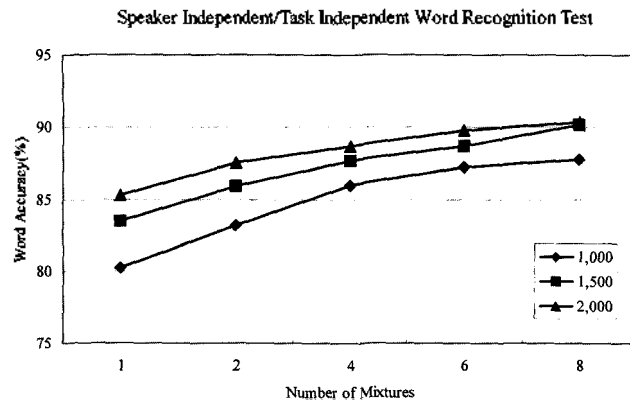


그림 7. 태스크 독립 KLE 452 여성 32명의 단어인식률  
 Fig. 7. Word accuracy of KLE452 female 32 speakers on task independent.

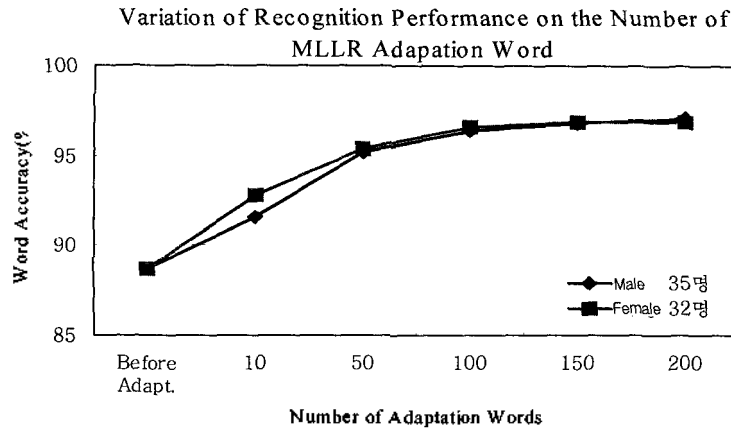


그림 8. 적응화 단어수에 따른 단어인식결과 변화  
 Fig. 8. Variation of word recognition performance on the number of adaptation word.

표 3. 452단어를 적응화한 후의 인식결과  
 Table 3. Recognition results after adapted by 452 words.

Exp.	Speakers	Adaptation	Recognition Results(%)
KLE452 Phone Recognition	Male 35 spks.	Before	22.8
		After	57.4
	Female 32spks.	Before	18.1
		After	56.0
KLE452 Word Recognition	Male 35 spks.	Before	88.7
		After	97.6
	Female 32spks.	Before	88.7
		After	98.0

표 4. 적응화 후의 연속음성인식 결과  
 Table 4. Sentence recognition results after adaptation.

Exp.	Adapt.	Adapt. Sent.	Grammar	Recognition Results (%)
AIR 200 Sentence Recognition (Male 14 spks.)	Before	-	2-gram	57.4
			3-gram	59.2
	After	10 sent. adapt.	2-gram	69.7
			3-gram	73.3
		100 sent. adapt.	2-gram	77.4
			3-gram	80.5

제안한 방법을 적용한 경우 적응화 전과 비교하여 각 화자의 10문장을 적응화한 경우 2-gram 문법에서 평균 12.3%, 3-gram 문법에서 평균 14.1%, 그리고 100문장을 적응화한 경우 2-gram 문법에서 평균 20.0%, 3-gram 문법에서 평균 21.3%의 연속음성인식을 향상을 각각 보였다. 연속음성인식의 경우 인식이 다소 저조한 원인은 학습에 사용된 음성과 평가에 사용된 음성의 발성환경이 서로 다르며, 적응화 학습에서 어느 정도 인식성능은 향상되었으나, 적응화에 사용된 연속음성 역시 다양한 문맥정보를 포함하고 있지 못함에 따른 것으로 사료된다.

이상의 실험결과로부터 화자독립 음향모델 학습에 사용된 음성 데이터의 양과 적응화 데이터의 양을 비교할 경우, 적응화 전의 화자독립 인식실험에서 매우 음소 및 단어인식률에서 낮은 인식률을 보이는데, 서론에서 지적한 것과 같이 모델의 학습에 많은 음성 데이터가 사용되었지만, 사용된 음성데이터의 문맥정보(음소환경)가 다양하지 못함을 나타내며 HM-Net으로 구성된 음성인식 시스템에서 특정 태스크에서의 인식성능은 우수하지만 태스크의 변화에 적절하게 대응하지 못하는 단점을 보여 주고 있다.

따라서 이를 해결하기 위해 본 논문에서는 적응화 방법에 의해 이 문제를 해결하고자 하였으며, 실험을 통하여 본 논문의 HM-Net 학습방법인 PDT-SSS 알고리즘의 문맥방향 상태분할을 개량하여 MLLR 적응방법의 회귀클래스 트리 생성에 적용한 결과 음소인식, 단어인식 및 연속음성인식 실험에서 일반적인 적응방법과 유사한 인식 성능을 보여 HM-Net 음성인식 시스템에서의 적용가능성을 보여주었으며, 그 유효성을 확인할 수 있었다.

## VI. 결론

본 논문에서는 HM-Net을 다양한 태스크에의 적용과 화자의 특성을 효과적으로 나타내기 위해 HM-Net 음성인식 시스템에 MLLR 적응방법을 도입하였으며, HM-Net 학습 알고리즘을 개량하여 회귀클래스 생성방법을 제안하였다. 제안방법은 PDT-SSS 알고리즘의 문맥방향 상태분할에 의한 상태레벨 공유를 이용한 방법으로, 문맥방향의 각 상태에 적응화자 음성데이터에 포함된 문맥정보를 분할하여 적응화될 음소환경을 결정하는 것이다. 따라서 제안방법은 새로운 화자로부터 문맥정보와 적응화 데이터의 발성 양에 의존하여 결정된 많은 적응 파라미터들을 (평균, 분산) 자유롭게 제어할 수 있게 된다. 제안방법의 유효성을 확인하기 위해 국어공학센터 452 음성 데이터와 항공편 예약관련 연속음성을 대상으로 인식 실험을 수행하였다. 그 결과, 전체적으로 음소인식의 경우 평균 34~37%, 단어인식의 경우 평균 9%, 연속음성인식의 경우 평균 20%의 인식성능 향상을 각각 보였다. 또한 적응화 데이터의 양에 따른 인식성능 비교에서 제안방법을 적용한 인식 시스템이 적응 데이터의 양이 적은 경우에도 향상된 인식성능을 보여 MLLR 적응방법의 특성을 만족하였다. 따라서 MLLR 적응방법을 도입한 HM-Net 음성인식 시스템에 제안한 회귀클래스 생성방법이 유효함을 확인할 수 있었으며, 향후에는 보다 다양한 태스크와 문맥정보가 많이 포함된 음성데이터를 이용하여 시스템의 유효성을 검토하고자 한다.

## 후 기

본 논문은 2000년도 한국과학재단 목적기초연구 (과제번호: 2000-1-30300-003-3) 지원으로 수행되었음.

## 감사의 글

본 연구가 수행되는데 지속적인 도움을 주신 일본 NIT 음성연구실의 Takaaki Hori 박사에게 감사드립니다. 그리고 본 연구에 사용된 음성데이터는 한국전자통신연구원과 국어공학센터에서 제공되었습니다.

## 참고 문헌

1. S. Sagayama and S. Honma, "Estimation of unknown context using a phoneme environment clustering algorithm," *Proc. of ICSP'90*, 1, 361-364, 1990.
2. K. F. Lee and H. W. Hon, "Large-vocabulary speaker independent continuous speech recognition using HMM," *Proc. of ICASSP'88*, 123-126, 1988.
3. J. Takami and S. Sagayama, "A successive state splitting algorithm for efficient allophone modeling," *Proc. of ICASSP'92*, 1, 573-576, 1992.
4. S. Young and P. Woodland, "The use of state tying in continuous speech recognition," *Proc. of EUROPEECH '93*, 2203-2206, 1993.
5. J. L. Gauvain and C. H. Lee, "Maximum a posteriori estimation for multivariate gaussian mixture observations of Markov Chains," *IEEE Trans. Speech Audio Processing*, 2, 291-298, 1994.
6. M. J. E. Gales and P. C. Woodland, Variance Compensation within the MLLR Framework, *Technical Report CUED/F-INFENG/TR242*, Cambridge University, 1996.
7. M. J. E. Gales, The Generation and Use of Regression Class Trees for MLLR Adaptation, *Technical Report CUED/F-INFENG/TR263*, Cambridge University, 1996.
8. R. Haeb-Umbach, "Automatic generation of phonetic regression class trees for MLLR adaptation," *IEEE Trans. on Speech and Audio Processing*, 9 (3), 299-302, 2001.
9. 오세진, 황철준, 김범국, 정호열, 정현열, "결정트리 상태 클러스터링에 의한 HM-Net 구조결정 알고리즘을 이용한 음성인식에 관한 연구," *한국음향학회지*, 21 (2), 199-210, 2002.
10. M. Ostendorf and H. Singer, "HMM topology design using maximum likelihood successive state splitting," *Computer Speech and Language*, 11, 17-41, 1997.
11. 임영춘, 오세진, 김광동, 노덕규, 송민규, 정현열, "음성인식에서 문맥의존 음향모델의 성능향상을 위한 유사음소단위에 관한 연구," *한국음향학회지*, 22 (5), 2003.
12. 성우창, 오세진, 김광동, 정호열, 정현열, "결정트리 상태 클러스터링에 의한 MLLR 적응화의 회귀 클래스 생성에 관한 연구," 2003년도 한국음향학회 하계학술발표대회 논문집, 22 (1s), 121-124, 2003.
13. T. Hori, M. Kaloh, A. Ito, and M. Kohda, "A study on HM-Nets using decision tree-based successive state splitting," *Proc. of ICSP'97*, 2, 383-387, 1997.
14. L. R. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993.
15. S.-J. Oh, K.-D. Kim, D.-G. Roh, W.-C. Sung, and H.-Y. Chung, "Speaker adaptation using regression classes generated by phonetic decision tree-based successive

state splitting," *Abstract Book of EUROSPEECH'03*, 51, 2003.

---

### 저자 약력

---

● 오 세 진 (Se-Jin Oh)

한국음향학회지 제22권 제5호 참조

● 성 우 창 (Woo-Chang Sung)



2002년 2월: 영남대학교 정보통신공학과 (공학사)  
2002년 3월 ~ 현재: 영남대학교 대학원 정보통신공  
학과 석사과정

※ 주관심분야: 음성분석 및 인식, 화자적응화, 디지털  
통신호처리

● 김 광 동 (Kwang-Dong Kim)

한국음향학회지 제22권 제5호 참조

● 노 덕 규 (Duk-Gyoo Roh)

한국음향학회지 제22권 제5호 참조

● 송 민 규 (Min-Gyu Song)

한국음향학회지 제22권 제5호 참조

● 정 현 열 (Hyun-Yeol Chung)

한국음향학회지 제22권 제5호 참조