

## 한국어 음성인식 성능향상을 위한 문맥의존 음향모델에 관한 연구

### A Study on Context-Dependent Acoustic Models to Improve the Performance of the Korean Speech Recognition

황철준, 오세진, 김범국, 정호열, 정현열

Chul-Joon Hwang, Se-Jin Oh, Bum-Koog Kim, Ho-Yeol Jung, Hyun-Yeol Chung

#### 요약

본 연구에서는 한국어 음성인식 성능향상을 위한 문맥의존 음향모델을 개선하기 위하여 한국어 음성학적 지식과 결정트리를 접목한 음소결정트리 기반 상태분할 알고리즘으로 한국어에 적합한 문맥의존 음향모델에 대해 고찰한다. HMM(Hidden Markov Model)의 각 상태를 네트워크로 연결하여 문맥의존 음향모델로 표현하는 HM-Net(Hidden Markov Network)이 있는데 이는 SSS(Successive State Splitting) 알고리즘으로 작성한다. 이 방법은 음향모델의 상태공유관계와 모델의 구조를 결정하는데 효율적이지만, 모델을 학습할 때 문맥환경에 따라 출현하지 않는 문맥이 존재하는 문제점이 있다. 본 연구에서는 이러한 문제점을 해결하기 위해 2진 결정트리와 SSS 알고리즘의 장점을 결합하여 문맥방향 상태분할을 수행할 때 각 노드에서 한국어 음성학적 지식으로 구성된 음소 길의어에 따라 상태분할 하는 방법으로서 PDT-SSS(Phonetic Decision Tree-based SSS) 알고리즘을 적용한다. 적용한 방법으로 작성한 문맥의존 음향모델의 유효성을 확인하기 위해 국어공학센터(KLE)의 452 단어와 항공편 예약관련 200문장(YNU 200)에 대해 화자독립 음소, 단어 및 연속음성인식 실험을 수행하였다. 인식실험결과, 문맥의존 음향모델에 대한 화자독립 음소, 단어 및 연속음성 인식실험에서 기존의 단일 HMM 모델보다 향상된 인식률을 보여, 한국어에 적합한 문맥의존 음향모델을 작성하는데 한국어 음성학적 지식과 음소결정트리 기반 상태분할 알고리즘이 유효함을 확인하였다.

#### ABSTRACT

In this paper, we investigate context-dependent acoustic models to improve the performance of the Korean speech recognition. The algorithms are using the Korean phonological rules and decision tree. By Successive State Splitting(SSS) algorithm, the Hidden Markov Network(HM-Net), which is an efficient representation of phoneme-context-dependent HMMs, can be generated automatically. SSS is a powerful technique to design topologies of tied-state HMMs, but it doesn't treat unknown contexts in the training phoneme contexts environment adequately. In addition it has some problem in the procedure of the contextual domain. In this paper, we adopt a new state-clustering algorithm of SSS, called Phonetic Decision Tree-based SSS (PDT-SSS), which includes contexts splits based on the Korean phonological rules. This method combines advantages of both the decision tree clustering and SSS, and can generate highly accurate HM-Net that can express any contexts. To verify the effectiveness of the adopted methods, the experiments are carried out using KLE 452 word database and YNU 200 sentence database. Through the Korean phoneme, word and sentence recognition experiments, we proved that the new state-clustering algorithm produce better phoneme, word and continuous speech recognition accuracy than the conventional HMMs.

**Key words:** HMMs, HM-Nets, SSS, PDT-SSS, Context-Dependent Acoustic Models, and Korean Phonological Rules

### I. 서론

현재 음성인식에서 널리 사용되고 있는 HMM(Hidden Markov Model)은 높은 인식성과 강건함, 대어휘, 화자 독립 음성인식 등과 같이 넓은 응용 분야에서 사용되고 있는 방법이다. 일반적으로 음성인식을 위한 인식단위로 확장성, 훈련성, 대용량성 등을 고려하여 유사음소단위(PLUs: Phone-Likely Units)가 많이 이용되고 있다[1]. 이 경우, 유사음소는 조음결합 등의 영향을 받기 때문에 하나의 모델로서 표현하는 데는 한계가 있다. 또한 모델의 파라미터 수가 많기 때문에 학습 데이터의 부족 문제가 다르게 되고 각 모델의 학습 데이터 수가 균일하지 못한 경우가 발생하게 된다.

최근에는 이상의 문제를 해결하기 위해 단일음소의 음향학적 특성을 변화시키는 환경요인으로 선행음소와 후행음소까지 고려한 문맥의존 모델로서 이음(allophone)을 인식 단위로 하는 방법이 소개되고 있으며 유효성이 확인되고 있다[2]. 문맥의존 모델은 음소의 평균 특징만을 가지며, 전후의 음소에 의존한 조음결합과 이음현상을 잘 표현할 수 있기 때문에 보다 정확한 음성인식을 가능하게 한다. 또한 문맥의존 모델의 유사한 상태의 확률분포를 공유하는 방법이 제안되고 있다[3]. 그러나 이음을 인식단위로 할 경우 단일음소의 경우와 비교하여 다음과 같은 두 가지 문제점을 해결할 필요가 있다. 첫째, 음소 환경에 의존하는 단위를 이용할 경우 총 모델의 수를 고려한 환경요인의 종류에 따라 모델의 수가 지수 함수적으로 증가하게 된다. 이는 학습 데이터가 충분하다면 문제가 없지만, 불충분할 경우 각 모델에 대응하는 학습 데이터의 부족으로 정확하게 모델의 파라미터를 추정할 수 없게 된다. 둘째, 학습 데이터가 유한하기 때문에 학습 데이터 중에 출현하지 않는 문맥이 존재할 수 있다.

이러한 문제점의 해결을 위해, 전자의 경우 비슷한 파라미터를 가지는 HMM의 상태와 출력확률분포를 하나로 하여 상향(bottom-up)으로 공유를 하는 방법이 제안되었다[3]. 그리고 모든 음소모델에 대응하는 상태공유를 자동으로 결정하는 SSS 알고리즘은 작은 상태에 보다 정확한 문맥의존 모델인 HM-Net을 자동적으로 생성할 수 있으며, 그 유효성이 입증되고 있다[4]. 후자의 경우 음소결정트리(Phonetic Decision Tree)에 의한 방법이 소개되고 있다[5]. 이 방법은 2진 트리 구조의 하향(top-down) 분할에 의해 음소간의 유사성에 기반한 음성학적 질의어를 통해 yes와 no의 분할을 수행하여 미지의 문맥을 작성하게 된다. SSS 알고리즘과 음소결정트리에 의한 상태 분할의 차이점은 SSS 알고리즘은 상태의 음향학적 분포확률 크기에 따라 상태분할을 결정하는 것이고, 음소결정트리는 특정 문맥정보에 따라 상태분할을 결정하는 것이다.

따라서, 본 연구에서는 한국어에 대해 보다 정확하게 미지의 문맥정보를 표현할 수 있는 문맥의존 음향모델을

작성하기 위해, SSS 알고리즘과 음소결정트리를 접목한 PDT-SSS(Phonetic Decision Tree based Successive State Splitting) 알고리즘을 적용한다. 적용한 방법의 유효성을 확인하기 위해 국어공학센터(KLE) 452단어와 본 연구실에서 남성 4인이 발성한 항공편 예약 관련 200문장을 대상으로 음소, 단어 및 문장 인식 실험을 통해 알고리즘의 유효성을 확인하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 한국어의 음성학적 규칙에 대하여 설명하고, 3장에서는 HM-Net을 구성하기 위한 SSS 알고리즘과 음소결정트리에 대하여 설명하고, 4장에서는 본 연구에서 이용하는 PDT-SSS 알고리즘을 설명한다. 5장에서는 적용된 방법을 이용한 음성인식 실험 및 고찰에 대하여 설명한 후, 마지막으로 6장에서 결론을 맺는다.

### II. 한국어 음성학적 규칙

한국어에는 다른 언어와는 달리 많은 문법과 음운규칙이 있다. 본 연구에서는 한국어에 적합한 문맥의존 음향 모델을 작성하기 위해 결정트리 기반 SSS 알고리즘의 상태분할에서 음소 질의어 집합의 구성에 한국어 음성학적 지식[6]을 이용하였다. 본 연구에서 적용한 음성학적 규칙을 표 1에 나타내었다.

표 1. 한국어 음성학적 규칙.  
Table 1. The Korean phonological rules.

		유성음		자	양순음		
		모 음			치(조)음		
모	혀위치	전	비원순	자	경구개음		
		설	원순		연구개음		
		중	비원순		성문음		
		설	원순		조음방법	파	무기연음
		후	비원순			열	유기경음
		설	원순			음	무기경음
	협(狹)		파	무기연음			
	반협(半狹)		찰	유기경음			
	광(廣)		음	무기경음			
	고모음		마	무기연음			
중고모음		찰	유기경음				
중저고음		음	무기경음				
저모음				비 음			
음	좁힘점위	경구개음		유 음			
		연구개음		반모음			
	인두음		목 음				
	좁힘점간	폐모음					
반폐모음							
개모음							

표 1에 나타낸 것과 같이 적용한 규칙은 크게 모음, 자음, 유성음, 비음, 유음, 반모음과 묵음으로 나뉜다. 이 중에서 모음은 혀의 위치, 입의 크기, 혀의 높이, 좁힘점위치, 좁힘점간극 등과 같이 크게 5부분으로 분류하였다. 그리고 자음은 조음자리, 조음방법 등과 같이 2부분으로 분류하고, 조음방법의 경우 파열음, 파찰음, 마찰음으로 다시 나누었다. 본 연구에는 음성학적 규칙을 문맥의 좌, 우를 포함하여 총 162부분으로 분류하였으며, 이를 이용하여 음소 질의어 집합을 작성하였다. 이렇게 작성한 음소 질의어 집합을 결정트리에 의한 상태분할에 사용하였다.

### III. SSS 알고리즘과 음소결정트리

#### 3.1 Hidden Markov Network(HM-Net)

SSS 알고리즘에 의해 작성한 HM-Net은 여러 개의 상태를 연결한 네트워크로 표현되며, HM-Net의 각 상태는 상태번호, 허용할 수 있는 문맥 클래스, 선행음소와 후행음소 리스트, 자기전이확률과 후행상태로의 전이확률 그리고 출력확률분포 파라미터와 같은 정보를 포함한다. HM-Net에서는 문맥정보가 주어지면, 이 문맥을 허용할 수 있는 상태를 선행상태와 후행상태 리스트의 제약 내에서 연결하여 이 문맥에 대한 모델을 결정할 수 있다. 이 모델은 자기전이와 다음 상태로의 전이만을 고려한 left-to-right 모델로 간주할 수 있으므로 일반적인 HMM과 같이 Baum-Welch 알고리즘에 의해 파라미터를 추정할 수 있다.

#### 3.2 SSS 알고리즘

SSS 알고리즘[4,7]은 모든 문맥을 나타내는 1상태의 초기모델로부터 문맥방향과 시간방향으로 상태분할 후 자동적으로 HM-Net[4,7]의 구조를 결정하는 알고리즘이다. SSS 알고리즘으로 HM-Net을 작성하는 단계를 그림 1에 나타내었다.

전체적으로 간략히 설명하면 다음과 같다. 우선 유사음소단위(PLUs)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 시간방향에 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.

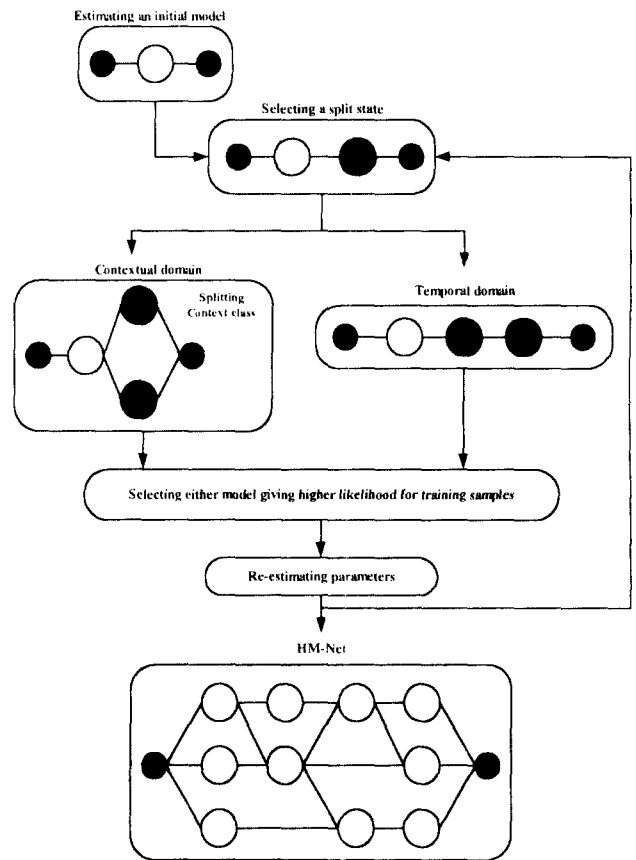


그림 1. SSS 알고리즘의 구성도.  
Fig. 1. Diagram of SSS algorithm.

#### 3.3 음소결정트리

음소결정트리[5]는 음소의 음향적 변동을 파악하는 것으로, 미지 음소환경의 음향적 특성을 예측하는 방법이다. 음소결정트리는 뿌리(root)를 음소환경에 독립한 2진 트리로 나타내고 뿌리에서 앞 방향으로 문맥클래스의 분할을 수행한다(그림 2). 이 트리는 뿌리에서 앞 방향으로 진행함에 따라 음소환경의 의존도가 강한 단위를 나타내는 계층적 구조를 가지며, 일반적으로 앞 부분에 모델을 대응시키게 된다. 트리의 각 노드에서는 경험적으로 음소 유사성에 기인한 질의어를 할당하여 yes와 no에 의해 문맥클래스를 두 개로 분할한다. 음소환경과 음소군에 따라서 각 질의어를 구성한다. 이러한 음소환경을 트리의 뿌리 노드에서 질의어를 찾아 반드시 앞에 대응시키기 위해, 미지의 음소환경에서 음향학적으로 가장 유사한 앞의 노드로 분류된다고 할 수 있다. 이를 위해, 출현하지 않는 음소환경을 음소환경 독립모델 등으로 대체할 필요도 있다.

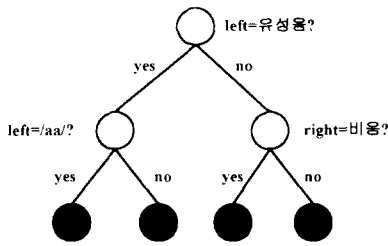


그림 2. 음소결정트리.  
Fig. 2. Phonetic-Decision tree.

#### IV. PDT-SSS 알고리즘

본 연구에서는 SSS 알고리즘에 한국어 음성학적 지식으로 구성된 음소질의어 집합을 이용한 음소결정트리에 기반한 상태분할 방법을 도입한 PDT-SSS(Phonetic Decision Tree-based SSS) 알고리즘[8]을 이용하였다. PDT-SSS는 SSS 알고리즘의 문맥방향 상태분할에 음소결정트리를 결합한 것으로 HM-Net에서 새로운 상태의 모델 파라미터 공유와 학습데이터에 출현하지 않는 미지의 문맥에 대한 학습을 수행할 수 있도록 구성되어 있다. 여기서 음소결정트리는 2진 트리로서 각 노드는 음소질의어로 구성되어 있다. 각 음소모델의 공유 파라미터는 각 트리의 잎(leaf) 노드와 연관되고, 문맥의존 모델은 음소질의어에 의해 트리의 뿌리(root) 노드에서 잎 노드까지 조사하여 임의의 문맥에 할당되어진다. PDT-SSS의 특징은 허용할 수 있는 문맥 클래스는 음소질의어에 따른 결정트리에 의해 분할된다는 것이다. 또한, 하나의 상태가 분할될 때, 두 개의 혼합수는 새로운 상태와 관련된 것이 아니고 새로운 상태에 대한 단일 가우스 분포는 학습 샘플로부터 계산된다. 따라서, PDT-SSS 알고리즘이 적절한 문맥 클래스의 분할과 임의의 문맥을 표현할 수 있기 때문에 보다 정확한 HM-Net을 작성할 수 있게 된다. PDT-SSS 알고리즘의 주요 내용은 다음과 같다.

- 1) 한국어 음성학적 지식에 의한 음소질의어 집합을 작성한다.
- 2) Baum-Welch 알고리즘으로 초기 HM-Net을 학습한다.(각 상태는 단일 가우스 분포)
- 3) SSS 알고리즘과 같이 식(1)에 의해 최적 분포를 가지는 상태를 선택한다.
- 4) 문맥방향과 시간방향으로 분할할 상태를 선택한다.
  - 각 음소질의어에 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 음소질의어에 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 정한 상태수에 도달할 때까지 단계 3부터 반복한다.

단계 3에서 분할될 상태의 선택은 식(1)에 의해 계산되어진다.

$$d_i = n_i \sum_{p=1}^P \frac{\sigma_{ip}^2}{\sigma_{Tp}^2} \quad (1)$$

여기서,  $\sigma_{ip}^2, \sigma_{Tp}^2$ 는 상태  $i$ 의 분포 분산과 모든 샘플의 분산(정규화 계수)을 나타내고,  $n_i$ 는 상태  $i$ 의 추정에 이용한 음소 샘플의 수를,  $P$ 는 특징 벡터의 차원 수를 각각 나타낸다.

#### V. 인식실험 및 고찰

한국어 음성학적 지식과 결정트리 기반 상태분할 알고리즘에 의해 작성한 한국어 문맥의존 음향모델의 유효성을 확인하기 위해 음소, 단어 및 연속음성 인식 실험을 수행하였다. 그림 3에 인식시스템의 전체 구성도를 나타내었다.

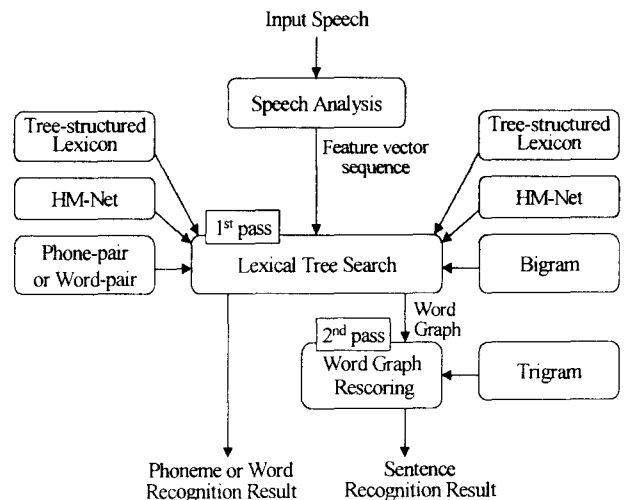


그림 3. 음성인식 시스템의 전체 구성도.  
Fig. 3. Overall diagram of speech recognition system.

음소 및 단어 인식실험에서는 문맥의존 음향모델을 작성하기 위해 사용된 음성데이터는 452단어를 38명이 2회 발성한 국어공학센터(KLE)의 음성 데이터베이스를 사용하였다. 이를 본 연구에서는 2 부분으로 나누어 학습과 평가에 사용하였다. 첫 번째 35명이 1회 발성한 15,820단어를 문맥의존 음향모델을 학습하는데 사용하였으며, 두 번째 학습에 참가하지 않은 3명이 첫 번째 발성한 1,356 단어를 화자독립 평가에 각각 사용하였다. 연속음성 인식 실험의 경우에는, 국어공학센터(KLE)의 단어음성과 본 연구실의 항공편 예약관련 200문장(YNU200) 연속음성 데이터베이스를 사용하였다. 음향모델의 학습을 위해 452 단어를 35명이 1회 발성한 15,820단어와 200문장을 8명이 1회 발성한 1,600문장을 문맥의존 음향모델을 학습하는데 사용하였으며, 학습에 참가하지 않은 4명의 200문장을 화자독립 연속음성인식 평가에 사용하였다.

모든 음성데이터는 16kHz의 샘플링과 16bits로 양자화되었으며,  $1-0.97z^{-1}$ 의 전달함수로 프리엠퍼시스 하였으며, 25ms의 해밍 윈도우를 곱하여 10ms씩 이동하면서 분석하였다. 이를 통해 음성 특징 파라미터는 12차 LPC-멜 캡스트럼 계수와 정규화된 대수 에너지에 1차 및 2차의 차분 성분을 포함하여 총 39차의 특징 파라미터를 구하였다. 또한 PDT-SSS 알고리즘에 의한 문맥방향의 상태 분할을 위해 162개(문맥의 좌, 우)의 음소 질의어 집합을 한국어 음성학적 지식에 근거하여 작성하였다. 초기 HM-Net의 구조는 48개의 유사음소단위를 병렬로 연결하여 141개의 상태를 가지도록 구성하였다. 모든 HM-Net은 혼합수 4를 가지며 200에서 1,200상태까지는 200상태씩 증가시켰으며, 상태수 2,000인 HM-Net도 학습하였다.

음소 및 단어인식 알고리즘은 One-Pass Viterbi beam 탐색 알고리즘[3,9]으로서 음소인식의 경우 한국어 음소 제약을 가지는 phone-pair 문법을, 단어인식의 경우 word-pair 문법을 각 사용하였다.

먼저 그림 4에 화자독립 음소인식 실험결과를, 그림 5에 화자독립 단어인식 실험결과를 각각 나타내었다.

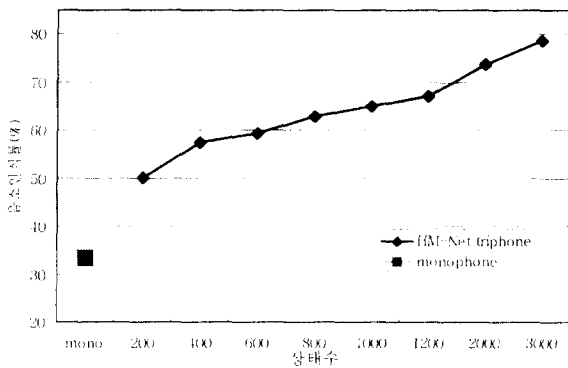


그림 4. 화자독립 음소인식 실험결과.

Fig. 4. Speaker-independent phoneme recognition results.

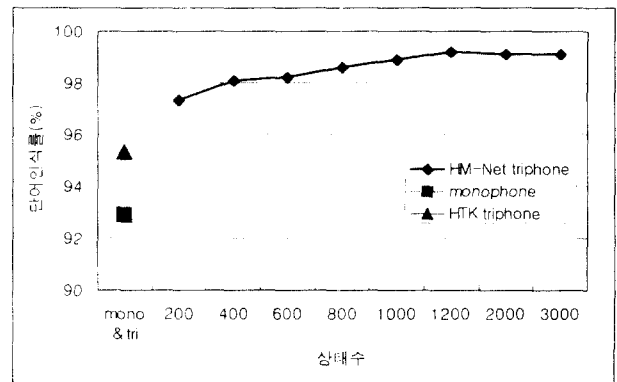


그림 5. 화자독립 단어인식 실험결과.

Fig. 5. Speaker-independent word recognition results.

그림 4의 음소인식의 경우, 모노폰에 대해 KLE 3인 평균 33.5%를 나타내고 있다. 그리고 HM-Net triphone에 대해서는 상태수 200일 때 평균 50.2%, 상태수 3,000일 때 평균 78.6%를 나타내고 있다. 모노폰과 상태수 3,000일 때의 HM-Net triphone을 비교하면 HM-Net을 이용한 경우가 평균 45.1%의 음소인식률 향상을 보였다. 또한 상태수 200과 상태수 3,000일 때의 HM-Net을 비교하면 평균 28.4%의 인식률 향상을 보이고 있다. 마찬가지로 그림 5의 단어인식률에서도 모노폰의 경우 KLE 3인 평균 92.9%, 상태수 200일 때 평균 97.3%, 상태수 3,000일 때 평균 99.1%의 평균 단어인식률을 구하였다. 그림 5에서도 모노폰과 HM-Net을 비교하면 상태수 3,000일 때의 HM-Net을 이용한 경우가 평균 6.2%의 인식률 향상을 나타내고 있다. 또한 본 연구에서 제안한 알고리즘의 유효성을 검토하기 위해 HM-Net 단어인식에서 사용한 동일한 화자와 한국어 질의어 집합을 이용하여 HTK에 의해 triphone 음향모델을 학습한 후 화자 3인에 대해 단어인식실험을 수행하였다. 그림 5에서 나타낸 것과 같이 본 연구에서 제안한 방법을 이용한 경우가 전체적으로 향상된 인식성능을 보였으며, 특히 HM-Net의 상태수가 1,200일 때 HTK를 이용한 경우보다 평균 3.9% 향상된 인식률을 얻었다.

그리고 본 연구에서는 한국어의 다양한 특성을 고려하여 48개의 유사음소단위(PLUs)를 사용하였는데, 만약 48개의 유사음소단위로 triphone을 작성한다면, 실제 음성인식 시스템에서 110,592( $48^3$ )개의 triphone을 만들어야 하지만 실제로 많은 수의 triphone을 작성하여 인식 시스템에 사용하면 계산적 부하가 발생한다. 실제 본 연구에서 학습에 사용된 음성 데이터에 출현하는 음소단위로 생성될 수 있는 triphone의 수는 2,164개이지만 PDT-SSS 알고리즘에 의한 문맥방향으로 한국어 음성학적 지식과 결정트리 기반 상태분할을 수행한 결과 108,289개의 HM-Net triphone을 작성할 수 있었고, 유사한 확률을 가지는 상태를 공유하여 시스템의 계산적 부하를 최소화하였다.

연속음성 인식 알고리즘은 Multi-pass 탐색 알고리즘 [3]으로서 1-pass 탐색의 경우, 단어 2-gram 언어모델을 이용하여 프레임 동기형 Viterbi beam 탐색을 수행한 후 단어 그래프를 출력한다. 2-pass 탐색의 경우 1-pass의 단어 그래프와 보다 정밀한 단어 3-gram을 이용하여 A\* stack decoding 탐색을 수행한 후 인식결과를 출력한다.

그림 6에 상태수의 변화에 따른 화자독립 연속음성인식률을 나타내고, 그림 7에 인식 문장에 포함된 단어인식률을 각각 나타내었다.

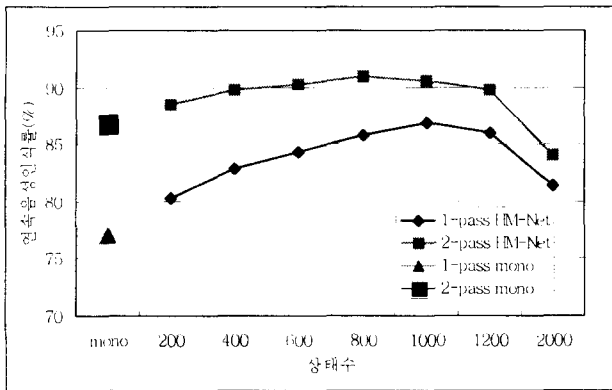


그림 6. 화자독립 연속음성인식률.

Fig. 6. Speaker-independent continuous speech recognition results.

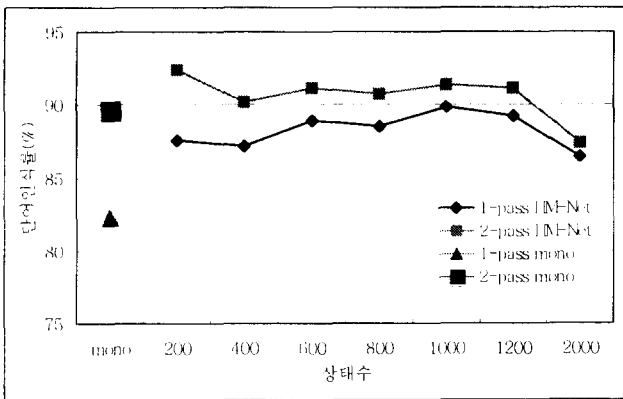


그림 7. 연속음성에 포함된 단어 인식률.

Fig. 7. Word recognition results included in continuous speech.

그림 6에서 상태수 1,000일 때 HM-Net triphone의 경우 1-pass의 인식률은 평균 86.9%로서 단일 HMM에 비해 평균 9.9%의 인식률을 향상을 보이고, 상태수 800일 때 HM-Net triphone의 경우 2-pass의 인식률은 평균 90.9%로서 단일 HMM에 비해 평균 4.1%의 인식률을 향상을 보였다. 또한 그림 7에서 인식대상인 연속음성에 포함된 798단어에 대한 인식률은 상태수 1,000일 때 HM-Net triphone의 경우 1-pass 인식률은 평균 89.9%로서 단일 HMM에 비해 평균 7.6%의 인식률 향상을 보이

고, 상태수 200일 때 HM-Net triphone의 경우 2-pass의 인식률은 평균 92.4%로서 단일 HMM에 비해 평균 2.8%의 향상된 인식률을 구하였다.

그리고 상태수의 증가에 따라 연속음성인식률과 단어 인식률이 감소하는 원인으로서는 학습에 참가한 음성 데이터의 부족으로 인해 정확한 HM-Net이 생성되지 못한 것으로 생각된다. 이는 향후 음향모델을 작성하는데 많은 양의 음성 데이터를 사용할 경우 해결할 수 있을 것으로 기대된다.

이상의 결과들로부터 본 연구에서 한국어에 적합한 문맥의존 음향모델을 작성하기 위해 적용한 한국어 음성학적 지식과 결정트리기반 상태분할 알고리즘의 유효성을 확인할 수 있었다.

## VI. 결론

본 논문에서는 한국어 음성인식 시스템의 성능향상을 위한 문맥의존 음향모델을 개선하기 위하여 한국어 음성학적 지식과 결정트리를 접목한 음소결정트리 기반 상태분할 알고리즘으로 한국어에 적합한 문맥의존 음향모델에 관하여 연구하였다. 적용 방법은 2진 결정트리와 SSS 알고리즘의 장점을 결합하여 문맥방향 상태분할을 수행할 때 각 노드에서 한국어 음성학적 지식으로 구성된 음소 질의어에 따라 상태분할 하는 PDT-SSS(Phonetic Decision Tree -based SSS) 알고리즘이다.

적용된 방법으로 작성한 문맥의존 음향모델의 유효성을 확인하기 위해 국어공학센터(KLE)의 452 단어와 항공편 예약관련 200문장(YNU 200)에 대해 화자독립 음소, 단어 및 연속음성인식 실험을 수행하였다. 인식실험결과, 문맥의존 음향모델에 대한 화자독립 음소, 단어 및 연속음성 인식실험에서 기존의 단일 HMM 모델보다 향상된 인식률을 보여, 한국어에 적합한 문맥의존 음향모델을 작성하는데 한국어 음성학적 지식과 음소결정트리 기반 상태분할 알고리즘이 유효함을 확인하였다.

접수일자 : 2001. 8. 14      수정완료 : 2001. 10. 22

본 논문은 1998년도 한국학술진흥재단 대학부설연구소과제(과제번호 98-005-E00017) 연구비에 의해 연구되었음

## 참고문헌

- [1] 김범국, 정현열, "가변장 음소모델을 이용한 음소인식," 한국음향학회지, 제16권, 제8호, 1997.
- [2] K.F. Lee, S. Hayamizu, H.W. Hon, C. Huang, J. Swartz, R. Weide, "Allophone Clustering for Continuous Speech Recognition," Proc. of

ICASSP'90, pp. 749-752, 1990.

[3] S.J. Young, P.C. Woodland, "State Clustering in hidden Markov model based Continuous Speech Recognition," Computer Speech and Language, Vol. 8, No. 4, pp. 369-383, 1994.

[4] J. Takamita, S. Sagayama, "A Successive State Splitting Algorithm for Efficient Allophone Modeling," Proc. of ICASSP'92, pp. 573-576, 1992.

[5] L.R. Bahl, P.V.de Souza, P.S. Gopalakrishnan, D. Nahamoo, M.A. Picheny, "Decision Trees for Phonological Rules in Continuous Speech," Proc. of ICASSP'91, pp. 185-188, 1991.

[6] 이호영, "국어음성학," 태학사, 1996.

[7] 오세진, 임영훈, 황진준, 김범국, 정현열, "Hidden Markov Network를 이용한 음향학적 음소모델 작성에 관한 검토," 2000년도 한국음향학회 학술발표대회 논문집, 제19권 제2(s)호, pp. 29-32, 2000. 11.

[8] S.J.Oh, C.J.Hwang, B.K.Kim, H.Y.Jung, H.Y.Chung, "A Study on Speech Recognition using New State Clustering Algorithm of HMM Net with Korean Phonological Rules," Proc. of IC AI'2001, U.S.A, 2001. 6.

[9] 中川聖一, "確率モデルによる音声認識," 日本電子情報通信学会, 1988.



김범국(Bum Koog Kim)

正會員

1990년 영남대학교 수학과(이학사)

1992년 영남대학교 전자공학과  
(공학석사)

1998년 영남대학교 전자공학과  
(공학박사)

1997년 3월 현재 대구과학대학 정보전자통신계열  
조교수

관심분야 : 음성분석 및 인식, 언어처리, 멀티모달 시스템



정호열(Ho Youl Jung)

正會員

1988년 아주대학교 전자공학과(공학사)

1990년 아주대학교 전자공학과  
(공학석사)

1993년 아주대학교 전자공학과  
(박사후료)

1998년 (포)리용국립응용과학원 전자공학전공(공학박사)

1998년 4월 1998년 12월 (포)CREATIS Post Doc

1999년 3월 현재 영남대학교 전자정보공학부 조교수

관심분야 : 음성·영상 신호처리, 인공지능,  
디지털 워터마킹 등



황진준(Chul Joon Hwang)

正會員

1996년 영남대학교 전자공학과(공학사)

1998년 영남대학교 전자공학과  
(공학석사)

2000년 영남대학교 전자공학과  
(박사후료)

2000년 3월 ~ 현재 대구과학대학 정보전자통신계열  
전임강사

관심분야 : 음성분석 및 인식, 디지털 신호처리



오세진(Se Jin Oh)

正會員

1996년 영남대학교 전자공학과(공학사)

1998년 영남대학교 전자공학과  
(공학석사)

1998년 현재 영남대학교 전자공학과  
(박사후료)

2001년 9월 현재 대구과학대학 정보전자통신계열  
전임강사

관심분야 : 음성분석 및 인식, 언어처리



정현열(Hyun Yeol Chung)

正會員

1975년 영남대학교 전자공학과(공학사)

1989년 일본 동북대학교 정보공학과  
(공학박사)

1989년 3월 현재 영남대학교  
전자정보공학부 교수

1992년 7월 1993년 7월 미국 CMU Robotics 연구소  
객원 연구원

1994년 12월 1995년 2월 일본 토요하시기술과학대학  
외국인 연구자

2000년 6월 2000년 8월 미국 Qualcomm Inc.  
수석 엔지니어

관심분야 : 음성인식, 화자인식, 음성합성 및  
DSP 응용분야