Hidden Markov Network 음성인식 시스템의 성능평가에 관한 연구

A Study on Performance Evaluation of Hidden Markov Network Speech Recognition System

오 세 진*, 김 광 동*, 노 덕 규*, 위 석 오*, 송 민 규*, 정 현 열**
Se-Jin Oh, Kwang-Dong Kim, Duk-Gyoo Roh, Seog-Oh Wi,
Min-Gyoo Song, Hyun-Yeol Chung

요 약

본 논문에서는 한국어 음성 데이터를 대상으로 HM-Net(Hidden Markov Network) 음성인식 시스템의 성능평가를 수행하였다. 음향모델 작성은 음성인식에서 널리 사용되고 있는 통계적인 모델링 방법인 HMM(Hidden Markov Model)을 개량한 HM-Net을 도입하였다. HM-Net은 기존의 SSS(Successive State Splitting) 알고리즘을 개량한 PDT(Phonetic Decision Tree)-SSS 알고리즘에 의해 문맥방향과 시간방향의 상태분할을 수행하여 생성되는데, 특히 문맥방향 상태분할의 경우 학습 음성데이터에 출현하지 않는 문맥정보를 효과적으로 표현하기 위해 음소결정트리를 채용하고 있으며, 시간방향 상태분할의 경우 학습 음성데이터에서 각 음소별 지속시간 정보를 효과적으로 표현하기 위한 상태분할을 수행하며, 마지막으로 파라미터의 공유를 통해 triphone 형태의 최적인 모델 네트워크를 작성하게 된다. 인식에 사용된 알고리즘은 음소 및 단어인식의 경우에는 One-Pass Viterbi 빔 탐색을 사용하며 트리 구조 형태의 사전과 phone/word-pair 문법을 채용하고 있다. 연속음성인식의 경우에는 단어 bigram과 단어 trigram 언어모델과 목구조 형태의 사전을 채용한 Multi-Pass 빔 탐색을 사용하고 있다. 전체적으로 본 논문에서는 다양한 조건에서 HM-Net 음성인식 시스템의 성능평가를 수행하였으며, 지금까지 소개된 음성인식 시스템과 비교하여 매우 우수한 인식성능을 보임을 실험을 통해 확인할 수 있었다.

ABSTRACT

In this paper, we carried out the performance evaluation of HM-Net(Hidden Markov Network) speech recognition system for Korean speech databases. We adopted to construct acoustic models using the HM-Nets modified by HMMs(Hidden Markov Models), which are widely used as the statistical modeling methods. HM-Nets are carried out the state splitting for contextual and temporal domain by PDT-SSS(Phonetic Decision Tree-based Successive State Splitting) algorithm, which is modified the original SSS algorithm. Especially it adopted the phonetic decision tree to effectively express the context information not appear in training speech data, on contextual domain state splitting. In case of temporal domain state splitting, to effectively represent information of each phoneme maintenance time the state splitting is carried out, and then the optimal model network of triphone types are constructed by tying the parameter. Speech recognition was performed using the one-pass Viterbi beam search algorithm with phone-pair/word-pair grammar for phoneme/word recognition, respectively and using the multi-pass search algorithm with n-gram language models for sentence recognition. The tree-structured lexicon was used in order to decrease the number of nodes by sharing the same prefixes among words. In this paper, the performance evaluation of HM-Net speech recognition system is carried out for various recognition conditions. Through the experiments, we verified that it has very superior recognition performance compared with the previous introduced recognition system.

Key words : Hidden Markov Network, Successive State Splitting, Phonetic Decision Tree-based SSS, State Splitting Mode, HM-Net Speech Recognition System

I. 서 론

음성은 인간에게 있어서 가장 자연스러운 정보전달 수단의 하나이다. 음성을 이용한 정보입력, 정보검색, 응용 프로그램의 작동 등을 실현할 수 있는 man-machine 인터페이스 기술은 고도 정보화 사회에 있어서 각종 정보기기의 유용성을 제고하는데 중요한 역할을 담당할 것으로 기대된다.

음성인식은 이와 같은 인터페이스를 구성할 수 있는 기초기술로서 지금까지 많은 연구가 수행되어오고 있다. 음성인식의

*한국천문연구원, **영남대학교 전자정보공학부
접수 일자 : 2003. 9. 05 수정 완료 : 2003. 10. 06
논문 번호 : 2003-4-7
※본 논문은 2000년도 한국과학재단 목적기초연구 (과제번호: R01-2000-000-00276-0) 지원으로 수행되었습니다.

최종 목표는 인간이 발성한 음성신호로부터 언어정보를 정확하게 추출하여 기계로 하여금 이해하도록 하는 것이다. 그러나 연속적으로 발성된 음성신호는 발성환경, 발성화자에 따라 변화가 많으므로 단어 및 음소와 같은 인식의 단위를 명확하게 구분하기 어렵고, 동일한 내용을 표현하는 방법도 여러 가지로 나타날 수 있으므로 컴퓨터가 이를 정확히 인식하고 이해하게 하는 것은 쉬운 일이 아니다. 지난 30여년 간 축적된 인식기술과 컴퓨터 성능 향상으로 발성화자, 발성내용, 발성방법 등에 제한을 둔 태스크를 대상으로 하는 음성인식 기술이 일부 실용화되고 있지만, 사람과 사람사이에서 이루어지는 자연스러운 발성에 의한 정보교환과 같은 man-machine 인터페이스 구현을 위해서는 보다 성능이 우수하고 적용범위가 넓은 강건한 음성인식 시스템의 개발이 필수적이라고 할 수 있다.

일반적으로 음성인식 시스템에서는 통계적인 방법으로 HMM(Hidden Markov Model)[1][2]을 널리 사용하고 있다. 즉, HMM은 음성인식을 위한 기본단위(음소)의 모델링과 인식을 수행하는 방법으로서 인식엔진으로 들어오는 음소와 음성인식 엔진내의 DB로 갖고 있는 음소를 결합해 단어와 문장을 만드는 방법을 말한다. 또한 HMM은 관측 불가능한 프로세스를 관측 가능한 다른 프로세스를 통해 추정하는 이중 확률처리 방법으로 현재 음성인식에 많이 사용되고 있다. 따라서 음성인식에서 HMM방식을 이용한다는 것은 음성인식의 최소단위(음소)를 모델링해 이를 이용해 음성인식 시스템을 구성하는 것을 말한다. 이에 따라 HMM의 장점은 다른 방법보다 인식률이 높다는 것이다. 그러나 현재의 HMM 훈련샘플이 충분하지 못할 경우에는 정확한 모델추정이 어려운 점과 음소문맥에 관한 지식이 필요하다는 등의 문제가 있다. HMM으로 인식 시스템을 구성할 경우 최소의 인식 단위가 되는 음소의 경우 monophone으로 구성하는 것과 하나의 음소를 기준으로 음소의 앞과 뒤의 영향을 고려한 triphone으로 구성할 수 있다. 여기서 monophone을 이용한 경우보다 triphone을 이용한 경우가 좀더 나은 인식성능을 보이고 있다. triphone의 경우 학습 샘플의 부족할 경우 학습이 제대로 수행되지 않아 인식성능을 저하시키지만, 이를 해결하기 위해 state-tying(음소모델의 각 음소의 상태에서 확률값이 유사한 정도에 따라 각 상태의 확률을 서로 공유하는 것)을 이용하고 있다.

HMM의 단점으로는 음소모델링을 위한 모델의 상태수를 결정할 때 기본적인 상태구조를 모델 학습이 종료할 때까지 기본구조를 가지게 된다. 하지만 사람의 음성은 기본적인 최소단위라고 하지만 각각의 발성길이가 서로 달라 기본적으로 주어진 모델의 구조로 모든 음성을 모델링할 경우 임의의 음소에 대해서는 부적절한 구조가 될 수 있으며, 이는 음성인식 시스템의 인식성능에 많은 영향을 미친다고 볼 수 있다.

따라서 본 논문은 현재 음성인식 분야에서 널리 사용되고 있는 HMM의 문제점을 해결하기 위해 개량된 HM-Net(Hidden Markov Network)[3][4][5]을 도입하여 한국어 음성데이터에 대해 음성인식 시스템의 성능평가를 수행하고자 한다. HM-Net은 HMM 음소모델을 네트워크 형식으로 표현한 것을 말한다. 즉, HMM의 각 상태를 임의의 노드로 설정하여 네트워크로 연결한 구조로 표현되며 기본적인 모델의 구조가 triphone HMM의 각 상태를 공유하게 된다. HM-Net은 모델의 기본구조를 결정할 때, HMM과 같이 모델 학습이 종료될 때까지 구조가 그대로 존재하는 것이 아니라 학습이 수행되는 동안 화자가 발성한 음성의 최소 기본단위의 길이(시간)에 따라 모델의 구조를 결정하므로 한 음소가 가지는 모델의 길이가 유동적이므로 HMM의 단점을 보완할 수 있게 된다. 도입한 HM-Net에 대해 국어공학센터(KLE), 한국전자통신연구원(ETRI) 그리고 항공편 예약(YNU)관련의 한국어 음성데이터를 대상으로 음소, 단어, 및 연속음성인식을 수행한 후 그 유효성을 확인하고자 한다.

본 논문의 구성은 다음과 같다. Ⅱ장에서는 모델 네트워크 구성방법으로 SSS 알고리즘과 이를 개량한 PDT-SSS 알고리즘에 대해 자세히 설명한다. Ⅲ장에서 HM-Net 음성인식 시스템과 사용된 인식알고리즘에 대해 간략히 기술한다. Ⅳ장에서는 사용한 음성 데이터 및 분석조건에 대해 설명하고 Ⅴ장에서는 HM-Net 음성인식 시스템의 유효성을 확인하기 위해 음소인식, 단어인식, 및 연속음성인식 실험을 수행하고 그 결과에 대해 고찰한 후, 마지막으로 Ⅵ장에서 본 논문의 결론을 맺는다.

II. 모델 네트워크 구성방법

2.1 SSS 알고리즘

SSS 알고리즘[3]은 모든 문맥을 나타내는 1 상태의 HM-Net으로부터 상태를 문맥방향과 시간방향으로 분할하여 자동적으로 HM-Net의 구조를 결정할 수 있는 방법이다. SSS에 의한 HM-Net의 생성은 5단계로 나눌 수 있으며, 이하 Sagayama 등[3]에서 제안한 원래의 기존 SSS에 대해서 소개한다.

단계 1: 초기 모델의 학습

초기모델은 하나의 상태에 출력분포가 단일 가우스 분포를 가지는 HMM을 이용하여, 전체 학습 샘플에 대해 학습한다.

단계 2: 분할할 상태의 결정

모든 상태 중에서 출력확률 분포가 가장 큰 상태를 분할할 상태로 선택한다. 출력확률 분포는 단일 가우스의 분산 값과 파라미터 추정에 이용한 학습샘플 수를 곱한 것을 기준으로 한다. i번째 상태에서 출력확률 분포의 크기 d_i는 식 (1)을 이용하여 계산한다.

$$d_i = n_i \times \sum_k^K \frac{\sigma_{ik}^2}{\sigma_{Tk}^2} \tag{1}$$

$$\sigma_{ik}^2 = \lambda_{i1}\sigma_{i1k}^2 + \lambda_{i2}\sigma_{i2k}^2 + \lambda_{i1}\lambda_{i2}(\mu_{i1k} - \mu_{i2k})^2$$

여기서, K는 파라미터의 차원, λ_{i1}, λ_{i2}는 상태 i의 가중계수, μ_{i1k}, μ_{i2k}는 상태 i의 k번째 평균, σ^2_{i1k}, σ^2_{i2k}는 상태 i의 k번째 분산, n_i는 상태 i일 때의 학습샘플의 수, 그리고 σ^2_{Tk}는 모든 학습샘플의 k번째 분산을 나타낸다.

단계 3: 상태 분할

분할을 위해 선택된 상태는 다시 두 단계로 분할할 수 있는데, 새로운 상태의 출력확률 분포를 아래와 같이 구하게 된다.

단계 3-1: 분할할 상태를 통해 전체 학습 샘플에 대해 Viterbi 알고리즘을 이용하여 상태가 출력하는 샘플의 부분 계열을 추출하게 된다.

단계 3-2: 단계 3-1에서 추출한 모든 학습 샘플의 부분계열을 이용하여 1상태, 혼합수 2의 HMM을 학습한다.

단계 3-3: 단계 3-2에서 구한 2개의 가우스 확률분포를 새롭게 분할할 상태에 각각 할당한다.

이와 같이 새로운 상태의 출력확률 분포를 구한 후, 새로운 상태의 위치를 시간방향(직렬)으로 연결한 경우의 학습 샘플에 대한 우도 P_t와 문맥방향(병렬)으로 연결한 경우의 우도 P_c를 계산하고 우도가 높은 쪽의 상태를 선택한다. P_t와 P_c는 아래와 같은 과정을 통해서 계산된다.

(a) 문맥방향(Contextual Domain)의 상태분할

문맥방향의 상태분할은 2가지 경로를 고려할 수 있는데 이때 각각의 학습샘플이 어느 쪽의 상태를 선택할 것인가를 결정할 필요가 있다. 각각의 학습샘플에 대해 문맥환경 요소(선행음소와 후행음소)로 그룹을 나누고 이 그룹에서 우도가 높은 상태를 결정하게 된다. 이때 상태의 결정은 식 (2)를 이용한다.

$$P_c = \max_j \sum_l \max\left(P_m(y_{jl}), P_M(y_{jl})\right) \qquad (2)$$

여기서, j는 현재 상태에서 문맥환경 요소를, y_{jl}은 요소 j의 값이 l번째 요소가 되는 학습 샘플의 부분집합을, $P_m(y_{jl})$는 y_{jl}을 상태 m에 할당할 때의 우도를, $P_M(y_{jl})$는 y_{jl}을 상태 M에 할당할 때의 우도를 각각 나타낸다. 문맥환경 요소 j를 결정한 후, 식 (3)을 이용하여 분포 e_{jl}을 결정하게 된다.

$$\begin{cases} e_{jl} \in \boldsymbol{E}_m, & (P_m(y_{jl}) \geq P_M(y_{jl})) \\ e_{jl} \in \boldsymbol{E}_M, & (P_m(y_{jl}) < P_M(y_{jl})) \end{cases} \qquad (3)$$

여기서, e_{jl}은 문맥환경 요소 j에 속하는 l번째 성분을, $\boldsymbol{E}_m$는 상태 m을 통과하는 성분들의 집합을, $\boldsymbol{E}_M$는 상태 M을 통과하는 성분들의 집합을 각각 나타낸다.

(b) 시간방향(Temporal Domain)의 상태분할

시간방향의 상태분할은 2개의 상태를 직렬로 연결하여 학습한다. 여기서는 상태의 위치에 따라 문맥방향의 상태분할과 마찬가지로 두 가지의 가능성이 존재하며 각각의 우도를 계산한 후 높은 우도를 가지는 상태의 우도 P_t를 선택한다.

단계 4: 분포의 재추정

문맥방향과 시간방향의 상태분할을 수행한 후의 새로운 상태는 단일 가우스 분포를 가지게 되는데, 모든 상태의 출력확률 분포가 혼합수 2의 가우스 분포로서 최적 파라미터를 가지도록 재학습하게 된다. 이후, 미리 정의한 상태수에 도달할 때까지 단계 2와 단계 3을 반복하여 수행하게 된다.

단계 5: 분포의 변화

지금까지의 단계를 통해서 HM-Net 모델의 각 상태는 혼합수 2의 가우스 확률분포를 가지게 되며, 최종적으로 HM-Net 모델이 각 상태마다 단일 가우스 출력확률 분포를 가지도록 HM-Net 전체를 재학습하게 된다.

2.2 기존 SSS의 문제점

기존 SSS 알고리즘의 문제점은 다음과 같이 두 가지로 요약할 수 있다. 첫 번째, 환경요인이외의 요인에 의해 2개의 혼합수를 가지는 분포로 분할될 가능성이 존재한다는 것이다. SSS에서 2개의 혼합분포는 문맥방향과 시간방향 등의 환경요인에 의해 분할되고 각 확률분포를 새로이 분할된 상태에 할당하게 된다. 따라서 이 확률분포에 의해 우도확률을 계산하고 분할할 방향과 문맥 클래스의 분할방법을 결정하게 된다. 그러나, 2개의 혼합분포는 특징 파라미터의 벡터공간에서 학습 데이터의 출현확률을 근사화하기 어려우며, 반드시 각 확률분포가 시간방향의 환경요인에 의해 나누어져야 하는 제한도 없다. 예를 들어, 환경요인으로서 선행음소와 후행음소를 고려할 경우, 2개의 혼합분포가 선행음소의 영향에 의해 나누어질 수도 있다. 따라서 정의한 환경요인이외의 요인에 의해 나누어진 확률분포를 이용할 수도 있으며, 상태의 분할이 잘 수행되지 않을 수도 있게 된다. 그리고 여러 명이 발성이 음성 데이터에 대해서 SSS를 수행한다면, 각 화자의 특성에 따라 나누어진 확률분포가 발생하기 때문에 적절한 HM-Net의 구조를 결정하기 어렵게 된다. 따라서 SSS에 의한 HM-Net의 구조결정은 일반적으로 여러 명의 화자로부터는 수행하지 않는다. 화자독립의 HM-Net을 생성하기 위해서는 대부분 화자 1명이 발성한 대량의 학습 데이터에 의해 HM-Net의 구조만을 결정하고 확률분포의 파라미터는 여러 명이 발성한 학습 데이터에 대해 재학습하는 방법을 이용한다. 그러나, 화자 1명이 발성한 대량의 음성 데이터를 이용함에도 불구하고 화자 1명의 HM-Net 구조가 모든 화자를 대상으로 한 HM-Net 구조에 대해서 적절하지 못한 점도 존재한다.

두 번째로는 HM-Net에서 허용할 수 없는 문맥정보가 다른 모델에서 사용될 수 있다는 것이다. SSS에서 문맥 클래스의 분할은 학습 데이터에 대한 누적 우도확률이 최대가 되는 상태를 분할하도록 결정한다. SSS에는 우도확률을 계산할 수 없는 미지 문맥을 어느 쪽의 상태에 할당할 것인가를 결정하는 것이 아니기 때문에 이 경우 출현하지 않는 미지 문맥은 삭제시킨다. 이 결과 HM-Net에는 허용할 수 없는 문맥이 생성되며, 이와 같은 문맥은 인식할 때 문맥독립 모델을 대신 사용한다. 일반적으로 문맥독립 모델을 대신 사용할 경우 인식성능이 좋지 못한 것으로 알려져 있다. 단, HM-Net은 각 상태에서 허용할 수 있

: 문맥 클래스를 환경요인으로 분할한 음소 클래스 집합으로 .(현하여 출현하지 않는 문맥을 허용할 수 있는 경우가 있는데 ' 를 문맥에 대한 보간이라고 하지만, 모든 문맥을 허용할 수 .다는 것은 아니다. 또한 상태수가 증가하면 문맥 클래스가 점 더 세분화되어 보정이 어렵게 된다.

2.3 개량된 PDT-SSS 알고리즘

PDT-SSS[6][7][8]는 기존 SSS 알고리즘의 문맥방향 상태 할에 음소결정트리를 결합한 것으로 HM-Net에서 새로운 상태의 모델 파라미터 공유와 학습데이터에 출현하지 않는 미 의 문맥에 대한 학습을 수행할 수 있도록 구성되어 있다. 특 기존 SSS 알고리즘을 개선시킨 점은 다음의 두 가지 로 요약할 수 있다.

① 각 상태는 단일 정규분포를 가지며, 상태를 분할할 때는 새로운 분포를 구한다.

② 문맥 클래스는 질의어에 의해 2개로 분할한다.

PDT-SSS는 출현하지 않는 문맥을 해결하기 위해 SS 알고리즘에 음소환경요소의 음소사이 거리를 계산하 는 방법을 도입하는 것을 고려할 수 있지만, 이 방법은 포의 분할에 관한 문제는 해결하지 못한다[3]. 그러나 개량된 PDT-SSS 알고리즘은 2개의 혼합분포를 새롭게 나누는 것이 아니고 질의어와 재추정에 의해 분포를 다 시 설정하는 것이다. 따라서 분포가 미리 정해진 환경요 이외의 요인에 의해 분할되지 않는다. 즉, 출현하지 않 는 문맥에 대한 문제와 분포의 분할문제를 동시에 해결 할 수 있다. 이하에 PDT-SSS 알고리즘에 대해 간략히 기술하였다.

단계 1: 초기모델의 학습

초기모델을 구성하고 파라미터를 학습한다. 초기모델의 구조는 임의로 결정할 수 있지만 각 상태의 출력 확률분 포는 단일 정규분포로 설정한다.

단계 2: 피 분할상태의 결정

분할할 상태를 식(1)에 의해 결정한다. 식(1)은 기존의 SSS 알고리즘과 마찬가지로 상태의 추정에 사용된 음소 샘플 수를 고려한 것과 같다.

단계 3: 상태의 분할

분할할 상태 $S(m)$을 분할한다. 우선, 새로운 분포 파 라미터를 구하기 위해 $S(m)$에 대응하는 학습 데이터를 Viterbi Segmentation[1]에 의해 추출한다. 다음으로 이하 에 나타낸 문맥방향과 시간방향의 상태 분할을 수행한다. 학습 데이터에 대한 우도확률이 큰 것을 다음의 분할을 의해 선택된다.

(a) 문맥방향의 상태분할

상태 $S(m)$에 허용할 수 있는 문맥 클래스를 $C(m)$이 라 하고, k번째의 요소를 c_k라고 한다. 추출할 때, 각 c_k에 대응하는 샘플의 평균 μ_k, 분산 σ_k^2, 총 프레임 수 f_k를 구하면 질의어 q에 의해 $C(m)$을 분할한 경우의 yes와 no의 분포 파라미터는 다음 식에 의해 구할 수 있 다.

$$\mu_{q,yes} = \frac{\sum_{c_k \in Q_q} f_k \mu_k}{\sum_{c_k \in Q_q} f_k} \tag{4}$$

$$\sigma^2_{q,yes} = \frac{\sum_{c_k \in Q_q} f_k \{\sigma_k^2 + (\mu_k - \mu_{q,yes})^2\}}{\sum_{c_k \in Q_q} f_k} \tag{5}$$

$$\mu_{q,no} = \frac{\sum_{c_k \notin Q_q} f_k \mu_k}{\sum_{c_k \notin Q_q} f_k} \tag{6}$$

$$\sigma^2_{q,no} = \frac{\sum_{c_k \notin Q_q} f_k \{\sigma_k^2 + (\mu_k - \mu_{q,no})^2\}}{\sum_{c_k \notin Q_q} f_k} \tag{7}$$

$\mu_{q,yes}$, $\sigma^2_{q,yes}$: $C(m)$을 질의어 q에 의해 분할할 때의 yes 측의 분포 평균과 분산

$\mu_{q,no}$, $\sigma^2_{q,no}$: $C(m)$을 질의어 q에 의해 분할할 때의 no 측의 분포 평균과 분산

Q_q: 질의어 q에 대해서 yes가 되는 문맥 클래스

$C(m)$을 질의어 q에 의해 분할하고, 동시에 yes와 no의 분포를 새로운 상태 $S'(m)$과 $S(M)$으로 할당한다. 그리고 $S(m)$을 통한 경로에서 표현된 학습 샘플 Y에 대한 대수 우도의 최대값 P_c를 식(8)에 의해 계산하고 P_c를 만족하는 질의어 q에 의해 분할된 HM-Net을 선택한다.

$$P_c = \max_q \left\{ \sum_{y \in Y_{q,yes}} P_m(y) + \sum_{y \in Y_{q,no}} P_M(y) \right\} \tag{8}$$

$Y_{q,yes}$: 질의어 q에 대해서 yes가 되는 Y의 부분 집합

$Y_{q,no}$: 질의어 q에 대해서 no가 되는 Y의 부분 집합

$P_m(y)$: $Y_{q,yes}$에 속하는 샘플 y를 $S'(m)$상의 경로에 할당한 경우의 대수 우도

$P_M(y)$; $Y_{q,no}$에 속하는 샘플 y를 $S(M)$상의 경로에 할당한 경우의 대수 우도

(b) 시간방향의 상태분할

$S(m)$을 시간방향으로 분할하고, 전에 위치하는 상태를 $S'(m)$, 후에 위치하는 상태를 $S(M)$이라 둔다. 이 경우에 허용할 수 있는 문맥과 분포 파라미터는 $S(m)$으로부터 그대로 복사한다. 이후 추출한 학습 샘플을 이용하여 $S'(m)$과 $S(M)$의 분포 파라미터만을 EM 알고리즘[1][2]에 의해 추정한다. 파라미터 재추정 식은 다음과 같다.

$$\hat{\mu}_s = \frac{\sum_n \sum_t \gamma_s(x_{nt})\, x_{nt}}{\sum_n \sum_t \gamma_s(x_{nt})} \tag{9}$$

$$\hat{\sigma}_s^2 = \frac{\sum_n \sum_t \gamma_s(x_{nt})\, x_{nt}^2}{\sum_n \sum_t \gamma_s(x_{nt})} - \hat{\mu}_s^2 \tag{10}$$

단, $s = S'(m)$, $S(M)$

μ_s, σ_s^2 : 상태 s의 분포 평균과 분산

x_{nt} : n번째에 추출된 샘플 x_n의 t번째 특징 벡터

$\gamma_s(x_{nt})$: 상태 $S'(m)$과 $S(M)$으로부터 x_n이 출력된 경우에 상태 s에서 x_{nt}가 관측될 확률

시간 방향으로 분할할 경우의 대수 우도 P_t에는 Y에 대한 대수 우도의 총합을 대입한다.

단계 4: 분포의 재추정

$P_c \geq P_t$라면 문맥방향으로 분할한 HM-Net, $P_c < P_t$라면 시간방향으로 분할한 HM-Net을 선택한다. $S'(m)$과 $S(M)$으로 M에 1을 추가하고 분할 영향 및 범위에서 파라미터를 재추정한다.

단계 5: 분포의 재추정 및 반복

미리 정의한 상태수에 도달하면 종료한다. 그렇지 않으면 단계 2에서 단계 4를 반복한다.

이상의 과정을 통해 HM-Net의 구조가 결정된다.

2.4 Hidden Markov Network[3][4][5]

HM-Net은 HMM 음소모델을 네트워크 형식으로 표현한 것을 말한다. 즉, HMM의 각 상태를 임의의 노드로 설정하여 네트워크로 연결한 구조로 표현되며 기본적인 모델의 구조가 triphone HMM의 각 상태를 공유하게 된다. 각 음소모델의 상태는 상태번호, 전후에 올 수 있는 문맥 클래스(class), 선행상태와 후행상태 리스트, 자기천이 확률과 상태천이 확률, 그리고 출력확률 분포 파라미터 등의 정보를 가지고 있다. HM-Net에서는 문맥정보가 주어질 경우 이 문맥을 만족하는 상태를 선행상태와 후행상태 리스트의 제약조건 내에서 서로 연결하여 이 문맥에 대한 모델을 하나로 결정할 수 있다. 여기서 설명한 모델은 자기천이와 이웃하는 상태로의 천이만을 허용하는 일반적인 left-to-right형의 HMM과 동일하다고 볼 수 있다. 하지만, 위에서 HMM의 단점으로 볼 수 있는 모델의 기본구조를 모델 학습이 종료될 때까지 그대로 가지는 것이 아니라 학습이 수행되는 동안 발성길이에 따른 음소의 시간에 따라 모델의 구조가 자동적으로 변하게 된다. 즉, 모음의 경우 상태수가 자음에 비해 좀 길어진 구조를 가지게 된다.

HM-Net의 단점으로는 HM-Net의 기본구조가 문맥의존 모델을 사용하기 때문에 모델의 학습에 관련된 음성을 인식할 경우에는 인식성능이 우수하지만(태스크 종속), 모델의 학습과 관련되지 않은 음성을 인식할 경우에는 인식성능이 다소 떨어지는 문제점(태스크 독립)이 있다. 이는 모델학습에 사용되지 않은 문맥정보가 포함될 경우에 이런 현상이 나타나게 된다. 따라서 이를 해결하기 위해서는 다양한 문맥정보가 포함된 많은 양의 음성 데이터를 이용하여 모델을 학습할 경우 어느 정도 이 문제점을 해결할 수 있다.

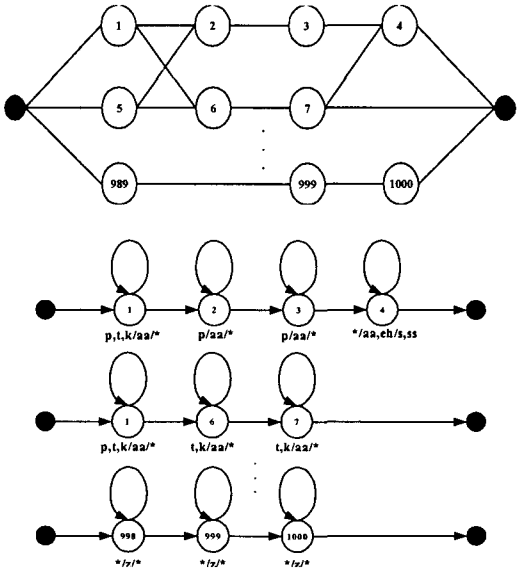

그림 1. HM-Net 모델의 예.

Fig. 1. An example of HM-Net models.

그림 1에 나타낸 HM-Net의 예에서 한 개의 음소를 전후해서 의존하는 triphone 모델의 경우 각 상태는 처리할 수 있는 문맥 클래스로서 선행/중심/후행 음소의 집합을 가지게 된다(단, "*"는 모든 음소의 집합을 나타냄). 그림 1에서 첫 번째 경로는 2개의 triphone(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 triphone(t/aa/*, k/aa/*)을 나타내는데, 여기서 "*"는 48개의 한국어 유사음소단위 중에서 임의의 하나를 의미한다.

Ⅲ. HM-Net 음성인식 시스템

그림 2에 HM-Net 인식 시스템의 전체 구성을 나타내었다. 이하에 인식 시스템의 핵심인 인식 알고리즘에 대해 간략히 설명한다.

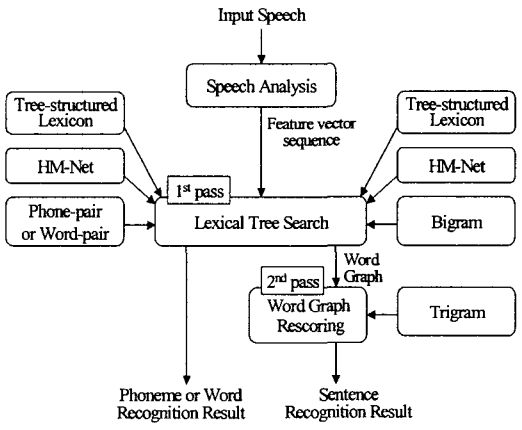

그림 2. HM-Net 음성인식 시스템의 전체 구성도.

Fig. 2. Overall schematic of HM-Net speech recognition system.

3.1 단계적 인식 방법

인식방법으로서 디코딩 알고리즘은 관측된 음성에 대해 가장 적절한 단어 후보열을 찾는 것이다. 인식 성능이 우수한 대어휘 연속음성인식을 위해서는 고차의 N-gram 언어모델[1]과 triphone과 같이 정밀한 음향모델을 이용하는 것이 바람직하다. 그러나 모델이 너무 정밀하면 인식 대상이 증가함에 따라 모델의 참조와 적용이 많아지므로 시스템에 많은 부하를 가져오게 된다. 따라서 1-pass 탐색만을 수행할 경우 전체적으로 처리 효율이 저하되므로 탐색 성능도 좋지 않다. 이를 해결하기 위해 탐색과정을 여러 개의 pass로 분할하여 간단한 모델에서 정밀한 모델까지 순서대로 적용하는 단계적 탐색법을 이용하고 있다[6]. 1-pass 탐색단계에서 입력음성 전체에 대해 간단한 모델을 이용하여 복잡한 인식처리를 수행하고 중간 결과를 출력한다. 그리고 2-pass 탐색단계에서는 1-pass 단계의 중간결과를 이용하여 탐색공간을 제한하여 정밀한 모델을 효율적으로 적용하여 재 탐색을 수행한다. 2-pass 탐색에서는 1-pass 탐색에서 입력음성 전체에 대해 미리 구해둔 예측 정보와 정밀한 모델을 이용하여 1-pass 탐색의 중간 결과보다 안정된 결과를 얻을 수 있다.

본 논문에서는 음소인식과 단어인식에서는 one-pass Viterbi 빔 탐색 알고리즘[1][2]으로서 음소인식의 경우 48개의 유사음소단위로 구성한 phone-pair 문법[1]을 사용하고, 단어인식의 경우 452단어로 구성한 word-pair 문법[1]을 사용하였다. 연속음성인식의 경우는 1-pass 탐색에는 단어 2-gram을 2-pass 탐색에는 단어 3-gram을 이용하는 multi-pass 탐색 알고리즘[6]을 이용하였다. 1-pass에서는 시스템의 고속화를 위해 프레임 동기형 빔 탐색 알고리즘을 이용하여 동적으로 목구조 형태의 사전의 각 상태에 2-gram 확률을 분할하여 지정한 후 사전에 있는 모든 단어에 대해서 탐색을 수행한다. 2-pass에서는 단어단위의 best-first의 스택 디코딩 탐색을 수행한다. 언어모델은 3-gram을 이용하고 단어단위의 탐색을 이용하는 것은 단어 레벨에서의 제약조건을 다루기 쉽고 정밀한 모델을 적용하는데 용이하기 때문이다. 1-pass에서의 중간 결과의 예측정보를 이용하기 위해 2-pass에서는 1-pass와 반대방향으로 탐색을 수행한다.

Ⅳ. 음성데이터 및 분석조건

KLE에서 제공된 452단어는 방음부스에서 채록되었으며, PBWs(Phoneme Balanced Words)로 구성되어 있다. 발성화자는 남성 38명과 여성 32명이 각각 2회씩 발성된 것으로 구성되어 있다.

ETRI에서 제공된 대어휘 음성DB는 성별, 연령별, 지역별로 분포된 1000명의 화자로 구성되어 있다. 남녀 화자의 성별은 50:50이며, 최소 SNR이 25dB이상인 조용한 사무실환경에서 수집되었다. 녹음장치는 일반 컴퓨터의 PC 마이크, PC 헤드셋, VoIP 통신망이 사용되었으며, VoIP 통신망은 PC 헤드셋으로 수집한 DB를 파일로 통신망에 전송하여 수집되었다. 총 발성 내용은 10,000 단어, 10,000 숫자음, 100,000 문장으로 구성되어 있다. 단어는 주식회사명, 지명, 인명, 상호명, 제품명, PC 명령어, PDA 명령어, 그 외 일반명사로 구성되어 있고, 숫자의 경우 번호독식(connected digit) 방식과 봉독식(natural number) 방식에 대해 수집되었다. 문장의 발성목록은 방송뉴스 대본에서 추출하였으며, 낭독체 50,000문장, 준낭독체 50,000문장으로 구성되어 있다.

또한 연속음성인식에 사용된 데이터는 항공편 예약관련 (YNU) 200 문장을 남성 12명이 1회 발성한 것으로 구성되어 있다.

본 논문에서는 KLE의 경우 남성 38명과 여성 32명의 1회 발성을 사용하였으며, ETRI의 경우 PC 헤드셋 마이크로 채록한 500명의 발성을 사용하였으며, YNU의 경우 8명 1회 발성은 학습, 나머지 4명은 평가에 사용하였다.

모든 음성 데이터는 16kHz, 16bits로 샘플링과 프리엠퍼시스 필터를 통과한 후 25ms의 해밍 윈도우를 곱하여 10ms씩 이동하면서 분석하였다. 이를 통해 음성 특징 파라미터는 12차 LPC-멜 켑스트럼 계수[1]와 정규화된 대수 에너지에 1차 및 2차의 차분 성분을 포함하여 총 26차와 39차의 특징 파라미터를 구하였다.

Ⅴ. 인식실험 및 고찰

KLE에서 제공된 단어음성 데이터와 ETRI에서 제공된 대어휘 음성DB 그리고 YNU의 항공편 예약 음성데이터를 이용하여 HM-Net 음성인식 시스템의 유효성을 확인하기 위해 음소 및 단어, 연속음성 인식실험을 각각 수행하였다. 실험은 다음과 같이 음향모델의 작성 방법에 따라 3가지로 구성된다. 우선 KLE의 남성 38명이 발성한 452 단어에 대한 인식 시스템 평가와 두 번째는 ETRI 대어휘 음성 데이터에 대한 인식 시스템 평가, 마지막으로 HM-Net의 모델 구조와 상태분할 모드에 따른 인식 시스템 성능평가로 나누어 수행되었다.

5.1 음향모델 작성

우선, KLE에서 제공한 452단어에 대한 기본적인 실험의 음향모델 작성은 남성 35명이 1회 발성한 452단어를 이용하였으며, 나머지 3명은 평가에 사용하였다. PDT-SSS 알고리즘의 문맥방향 상태분할을 위해 162개(문맥의 좌, 우)의 음소 질의어를 한국어 음성학적 지식에 근거하여 작성하였다. 초기 HM-Net의 구조는 48개의 유사음소단위를 병렬로 연결하여 141개의 상태를 가지도록 구성하였다. HM-Net 모델은 26차와 39차의 특징 파라미터를 사용하여 혼합수 4를 가지며 200에서 1200상태까지는 200상태씩 증가시켰으며, 상태수 2000인 HM-Net도 학습하였다.

두 번째로 ETRI에서 제공한 대어휘 음성데이터를 이용한 음향모델작성은 남녀 각 200명이 발성한 280발성을 이용하였

으며, 나머지 남녀 각 25명의 100단어를 평가에 사용하였다. 남녀 각 200명이 발성한 데이터에서 학습에 사용된 총 단어는 5,287단어이고 평가에 사용된 총 단어 중 학습에 포함된 단어 수는 607단어로서 약 11.5%의 단어가 중복 사용되었다. ETRI 데이터의 경우 상태분할을 위해 152개의 음소 질의어를 준비하였으며, 초기 HM-Net의 구조는 43개의 유사음소단위를 병렬로 연결하여 126개의 상태를 가지도록 구성하였다. HM-Net 모델은 39차의 특징 파라미터를 사용하여 혼합수 1, 2, 4, 6, 8개를 가는 상태수 1000, 1500, 2000인 모델을 각각 학습하였다.

마지막으로 KLE/YNU의 음성데이터를 이용한 음향모델 학습에는 KLE 452 단어를 남성 35명이 1회 발성한 15,820단어와 YNU 200문장을 남성 8명이 1회 발성한 1,600문장을 함께 사용하였다. 또한 학습에 참가하지 않은 KLE의 남성 3명의 452단어의 경우 화자독립 음소와 단어인식 평가에 사용하였으며, YNU의 남성 4명의 200문장을 화자독립 연속음성인식 평가에 사용하였다. 여기서는 첫 번째 실험을 위한 음향모델 작성과 마찬가지로 초기 HM-Net의 구조는 48개의 유사음소단위를 병렬로 연결하여 141개의 상태를 가지도록 구성하고, 39차의 특징 파라미터를 사용하여 혼합수 1, 2, 4를 가지며 200에서 1200상태까지는 200상태씩 증가시켰으며, 상태수 2000인 HM-Net도 학습하였다.

5.2 KLE 452 음소/단어인식 실험

KLE 단어음성 데이터에 대해 HM-Net 음성인식 시스템 기본적인 성능평가를 위해 남성 3명이 발성한 452단어의 첫 번째 발성에 대해 phone-pair/word-pair 문법을 가진 one-pass Viterbi 빔 탐색을 이용하여 음소 및 단어인식실험을 수행하였다.

표 1. HM-Net 상태수 및 차원별 음소/단어인식률.
Table 1. Phoneme/word recognition accuracy for HM-Net states number and feature parameter order.

차원	실험	HM-Net 상태수						
		200	400	600	800	1000	1200	2000
26차	음소	46.7	53.6	56.0	57.6	59.8	61.8	68.6
	단어	96.2	97.9	98.8	99.0	99.2	98.9	99.6
39차	음소	50.0	57.7	60.7	63.2	66.1	68.6	75.2
	단어	97.2	98.4	98.5	98.7	99.0	99.0	99.2

표 1에 나타낸 것과 같이 음소/단어인식 모두 각 차원에 대해 상태수가 증가할수록 인식률이 향상되는 것을 볼 수 있다. 음소인식의 경우 26차원에 비해 39차원의 경우가 상태수 2000개인 모델에서 평균 6.6% 향상되었으나, 단어인식의 경우 상태수 2000개인 모델에서 두 차원에 대해 높은 인식률을 보이지만 39차원이 26차원에 비해 평균 0.4% 낮은 결과를 보였다. 이는 학습 데이터가 적은 경우 26차원이 보다 효과적인 것으로 생각된다. 즉, 적은 학습 데이터에 대해 상태수를 계속해서 증가시킬 경우 학습 데이터의 부족으로 인해 모델 학습이 제대로 수행되지 않은 원인도 된다.

5.3 ETRI/KLE 데이터에 대한 인식 실험

ETRI 음성 데이터로 작성한 각 HM-Net 모델에 대해 2099개의 인식단어 카테고리 내에서 ETRI의 남녀 각 25명이 발성한 100단어와 KLE의 남성 35명과 여성 32명의 452단어를 대상으로 4.2절과 동일한 인식방법으로 태스크종속 및 태스크 독립 단어인식 실험을 수행하였다. 또한 ETRI의 남녀 각 25명이 발성한 60문장에 대해 단어 bigram 문법과 동일한 인식방법에 대해 연속음성인식 실험을 수행하였으며, 그 결과를 그림 3, 4, 5와 표 2에 각각 나타내었다.

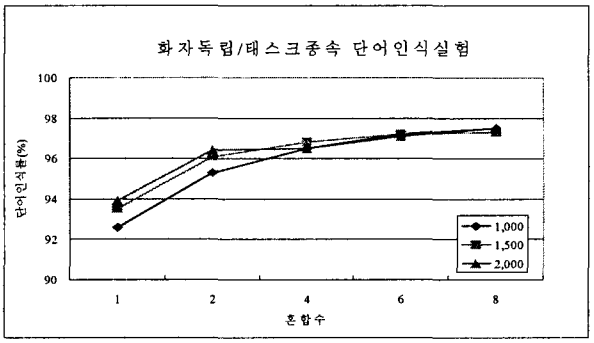

그림 3. 태스크 종속 ETRI 남녀 각 25명의 단어인식률.
Fig. 3. Word recognition accuracy of ETRI each 25 males and females for task dependent.

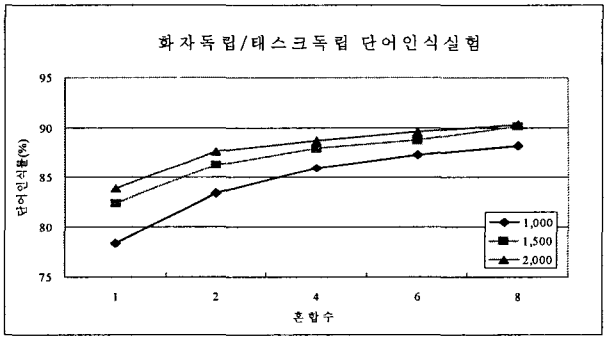

그림 4. 태스크 독립 KLE 452 남성 35명의 단어인식률.
Fig. 4. Word recognition accuracy of KLE 452 35 males for task independent.

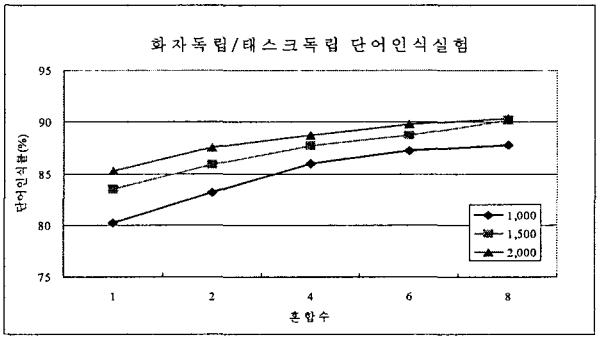

그림 5. 태스크 독립 KLE 452 여성 32명의 단어인식률.
Fig. 5. Word recognition accuracy of KLE 452 32 females for task independent.

표 2. 남녀 각 25명의 60문장 인식률.
Table 2. 60 sentences recognition accuracy of each 25 males and females.

상태	실험	혼 합 수(mixture)	
		4	8
1000	문장	95.5	96.7
	단어	97.7	98.3
1500	문장	95.5	96.7
	단어	97.6	98.3
2000	문장	94.9	96.6
	단어	97.0	98.2
mono	문장	94.5	
	단어	97.3	

그림 3의 태스크 종속 단어인식률의 경우, 학습 데이터의 양에 비해 모델의 상태분할 및 학습이 충분하지 않지만, 평균 97%를 상회하는 결과를 보이고 있다. 하지만 그림 4, 5에 나타낸 태스크 독립 단어인식의 경우 태스크 종속에 비해 인식률이 저조한데, 이는 HM-Net 모델을 학습할 때 많은 문맥환경과 발성리스트를 공유할 경우 높은 인식성능을 보이나 그렇지 못한 경우에는 다소 인식성능이 저하되는 것으로 생각된다. 즉, 태스크 독립실험의 경우, 모델의 상태분할과 학습에 사용한 음성 데이터의 문맥정보와 인식에 사용된 음성 데이터의 문맥정보가 서로 상이한 것과 포함되지 않은 것이 많이 있어 예상보다 저조한 인식률을 보이는 것으로 생각된다. 또한 사용된 음성 데이터의 문맥정보를 분석해 보면, ETRI 음성 데이터의 triphone 수는 16,380개로 문맥정보는 많지만, KLE 452단어 (triphone 수: 2,164개)에서의 triphone을 만족할 만큼 충분히 표현하지 못하는 것으로 생각된다. 그리고 ETRI 데이터의 경우 데이터 양은 많지만 다양한 문맥정보(음소)가 KLE의 PBW와 비교하여 균형있게 분포하지 못하고, 문맥정보에 대한 음성 데이터가 적은 것도 인식률 저하의 원인으로 생각된다.

표 2의 연속음성인식 실험결과의 경우 태스크 종속으로 비교적 높은 인식률을 보이고 있다. 하지만 상태수와 혼합수가 증가함에도 인식률이 조금 저하되는 경향이 있는데 이는 앞에서 설명한 것과 같이 많은 문맥정보(음소)에 대한 음성 데이터의 양이 부족한 것에 기인한 것으로 생각된다. 이는 향후 다양한 문맥정보(음소)에 대해 많은 음성 데이터를 사용할 경우 해결될 것으로 생각된다.

5.4 KLE/YNU 데이터에 대한 음소/단어/연속음성인식 실험

여기서는 PDT-SSS 알고리즘의 두 가지 상태분할 방법(시간방향, 문맥방향)을 선택하기 전, 이 두 가지 분할 모드를 선택하는 frame과 divergence에 의한 상태분할 모드에 따른 인식성능을 비교 고찰하고자 한다. 음소/단어인식에 사용된 인식방법과 문법은 첫 번째 실험과 동일한 방법을 사용하였으며, 연속음성인식은 두 번째 실험과 동일한 방법을 사용하였다.

우선, 그림 6, 7, 8에 상태수의 변화와 혼합수의 변화, 그리고 상태분할 모드를 divergence로 한 경우의 화자독립 음소, 단어 및 연속음성인식률을 나타내고, 그림 9, 10, 11에는 상태분할 모드를 frame으로 한 경우와 divergence로 한 경우의 음소, 단어 및 연속음성인식률의 비교를 각각 나타내었다.

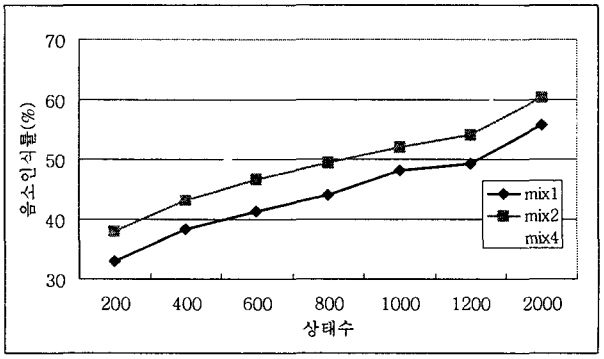

그림 6. Divergence 모드, 상태수, 혼합수에 따른 음소인식률 변화.
Fig. 6. Phoneme recognition accuracy variation according to divergence mode, state number, mixture number.

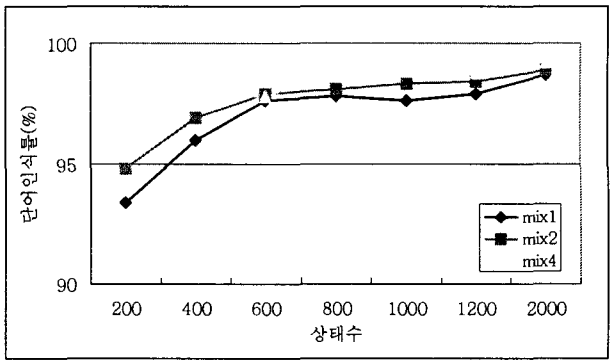

그림 7. Divergence 모드, 상태수, 혼합수에 따른 단어인식률 변화.
Fig. 7. Word recognition accuracy variation according to divergence mode, state number, mixture number.

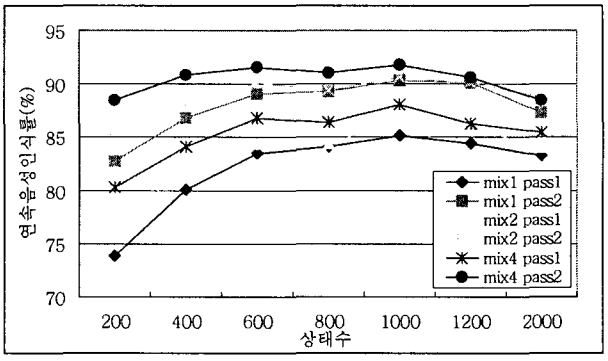

그림 8. Divergence 모드, 상태수, 혼합수에 따른 연속음성인식률 변화.
Fig. 8. Sentence recognition accuracy variation according to divergence mode, state number, mixture number.

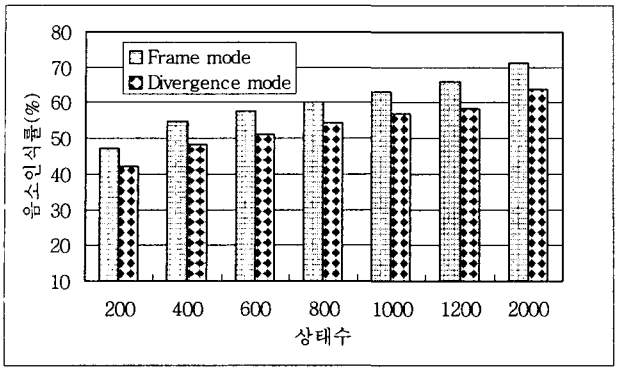

그림 9. Frame과 Divegence 상태분할 모드에 따른 음소인식률의 비교(혼합수 4).

Fig. 9. Phoneme recognition accuracy comparison according to frame and divergence state splitting mode(mixture 4).

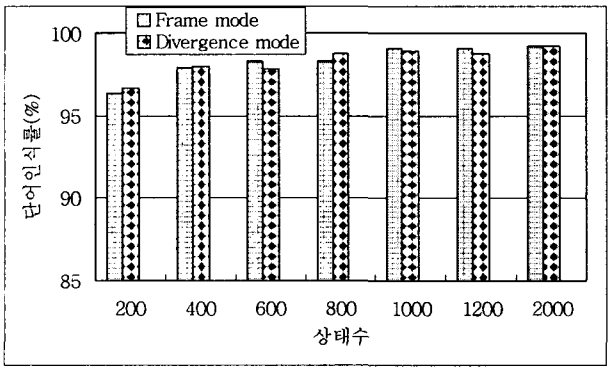

그림 10. Frame과 Divegence 상태분할 모드에 따른 단어인식률의 비교(혼합수 4).

Fig. 10. Word recognition accuracy comparison according to frame and divergence state splitting mode(mixture 4).

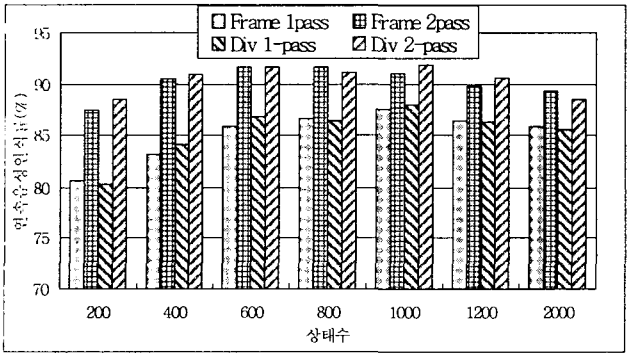

그림 11. Frame과 Divegence 상태분할 모드에 따른 연속음성인식률의 비교(혼합수 4).

Fig. 11. Sentence recognition accuracy comparison according to frame and divergence state splitting mode(mixture 4).

그림 6, 7, 8의 경우에서 혼합수와 상태수가 증가할수록 인식률이 증가하는 것을 볼 수 있으나, 그림 8의 경우 1000상태 이후에 인식률이 감소하는 것을 볼 수 있다. 이는 모델학습에 사용된 음성 데이터가 부족하여 이런 결과를 초래한 것으로 생각된다. 그리고 그림 9, 10, 11의 비교에서 전체적으로는 모델의 상태수가 증가할수록 frame에 기반한 분할할 상태 선택방법에 의한 인식결과가 우수하지만 모델의 상태수가 적은 경우에는 divergence에 의한 분할할 상태 선택방법이 조금 더 나은 인식결과를 보였다. 따라서 인식 시스템을 구축할 때 상태분할 모드의 선택과 최적의 상태수 그리고 혼합수의 결정은 인식률에 매우 많은 영향을 미침을 확인할 수 있었다.

Ⅵ. 결 론

본 논문에서는 한국어 음성 데이터를 대상으로 HM-Net 음성인식 시스템의 성능평가를 수행하였다. 음향모델 작성은 음성인식에서 널리 사용되고 있는 통계적인 모델링 방법인 HMM을 개량한 HM-Net을 도입하였다. HM-Net은 원래의 SSS 알고리즘을 개량한 PDT-SSS 알고리즘에 의해 문맥방향과 시간방향의 상태분할을 수행하여 생성되는데, 특히 문맥방향 상태분할의 경우 학습 음성데이터에 출현하지 않는 문맥정보를 효과적으로 표현하기 위해 음소결정트리를 채용하고 있으며, 시간방향 상태분할의 경우 학습 음성데이터에서 각 음소별 지속시간 정보를 효과적으로 표현하기 위한 상태분할을 수행한 후, 파라미터의 공유를 통해 triphone 형태의 최적인 모델 네트워크를 작성하게 된다. 인식에 사용된 알고리즘은 음소 및 단어인식의 경우에는 One-Pass Viterbi 빔탐색을 사용하며 목구조 형태의 사전과 phone/word-pair 문법을 채용하고 있다. 연속음성인식의 경우에는 단어 bigram과 단어 trigram 언어모델과 목구조 형태의 사전을 채용한 Multi-Pass 빔탐색을 사용하고 있다. 전체적으로 본 논문에서는 다양한 조건에서 HM-Net 음성인식 시스템의 성능평가를 수행하였으며, 지금까지 소개된 음성인식 시스템과 비교하여 매우 우수한 인식성능을 보임을 실험을 통해 확인할 수 있었다.

※본 논문에서 사용된 음성 데이터는 한국전자통신연구원과 국어공학센터에서 제공되었습니다.

참 고 문 헌

[1] L. R. Rabiner, and B. H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993.

[2] 中川聖一, *確率モデルによる音聲認識*, 日本電子情報通信學會, 1988.

[3] J. Takami, and S. Sagayama, "A successive state splitting algorithm for efficient allophone modeling," *Proc. of ICASSP'92*, Vol. 1, pp. 573-576, 1992.

[4] M. Suzuki, S. Makino, A. Ito, H. Aso, and H. Shimodaira, "A new HMnet construction algorithm requiring no contextual factors," *IEICE Trans. Info. & Syst.*, Vol. E78-D, No. 6, pp. 662-669, 1995.

[5] M. Ostendoft, and H. Singer, "HMM topology

design using maximum likelihood successive state splitting," *Computer Speech and Language,* Vol. 11, pp. 17-41, 1997.

[6] T. Hori, *A study on large vocabulary continuous speech recognition,* Ph. D. thesis, Yamagata University, Japan, 1999.

[7] Se-Jin Oh, Cheol-Jun Hwang, Bum-Koog Kim, Hyun-Yeol Chung, and Akinori Ito, "New state clustering of hidden Markov network with Korean phonological rules for speech recognition," *IEEE 4th workshop on Multimedia Signal Processing*, pp. 39-44, 2001.

[8] 오세진, 황철준, 김범국, 정호열, 정현열, "결정트리 상태 클러스트링에 의한 HM-Net 구조결정 알고리즘을 이용한 음성인식에 관한 연구," *한국음향학회지*, 제21권 제2호, pp. 199-210, 2002.

[10] Se-Jin Oh, Cheol-Jun Hwang, Bum-Koog Kim, Hyun-Yeol Chung, "Performance Evaluation of HM-Nets Speech Recognition System using the Large Vocabulary Korean Speech Databases," *Proc. of Kyushu-Youngnam Joint Conference on Acoustics*, pp. 49-52, Japan, 1. 2003.

[11] 오세진, 김광동, 노덕규, 송민규, 황철준, 김범국, 정현열 "한국어 대어휘 음성 DB를 이용한 HM-Net 음성인식 시스템의 성능평가," *2003년도 대한전자공학회 하계종합학술발표대회 논문집 IV,* 제26권 제1호, pp. 2443-2446, 7. 2003.

[12] 오세진, 김광동, 노덕규, 송민규, 황철준, 김범국, 성우창, 정현열, "상태분할 모드에 따른 HM-Net 음성인식 시스템의 성능평가," *제 16회 신호처리합동학술대회 논문집*, 강원대 9. 2003.

오 세 진(Se-Jin Oh)
正會員
1996년 2월 영남대 전자공학과(공학사)
1998년 2월 영남대 전자공학과(공학석사)
2002년 2월 영남대 전자공학과(공학박사)
2001년 9월 ~ 2002년 12월 대구과학대학 전임강사
2002년 12월 ~ 현재 한국천문연구원 선임연구원
관심분야 : 디지털신호처리, 음성처리, DSP 응용

김 광 동(Kwang-Dong Kim)
正會員
1973년 2월 영남대 전기공학과(공학사)
1975년 5월 ~ 1982년 8월 고미반도체(주) 기술과장
1982년 9월 ~ 1986년 7월 대한통운(주) 전산실장
1986년 9월 ~ 1993년 4월 제성전자(주) 기술부장
1993년 4월 ~ 현재 한국천문연구원 책임연구원
관심분야 : 디지털신호처리, 상관기, DSP 응용분야

노 덕 규(Duk-Gyoo Roh)
正會員
1985년 2월 서울대학교 천문학과(이학사)
1994년 8월 일본 동경대학 대학원 이학계 연구과 천문학전공(이학석사)
2002년 2월 일본 동경대학 대학원 이학계 연구과 천문학전공(박사수료)
1984년 4월 ~ 현재 한국천문연구원 선임연구원
관심분야 : 디지털신호처리, DSP 응용 분야

위 석 오(Seog-Oh Wi)
正會員
1993년 2월 전남대 전기공학과(공학사)
1996년 2월 전남대 전기공학과(공학석사)
2002년 8월 전남대 전기공학과(공학박사)
2003년 8월 ~ 현재 한국천문연구원 선임연구원
관심분야 : 안테나 제어, 전력전자, DSP 응용 분야

송 민 규(Min-Gyu Song)
正會員
2001년 2월 강원대 전기공학과(공학사)
2003년 2월 강원대 전자공학과(공학석사)
2002년 12월 ~ 현재 한국천문연구원 연구원
관심분야 : 디지털신호처리, DSP 응용 분야

정 현 열(Hyun-Yeol Chung)
正會員
1975년 2월 영남대학교 전자공학과(공학사)
1989년 3월 일본 동북대학교 정보공학과(공학박사)
1989년 3월 ~ 현재 영남대학교 전자정보공학부 교수
관심분야 : 음성인식, 화자인식, 음성합성, DSP 응용