울산대학교 공학연구논문집 제30권 2호 pp. 333~358, 울산대학교.
Journal of Engineering Research Vol. 30, No. 2, pp. 333~358, 1999
University of Ulsan.

한국어 음소의 HMM 모형에 관한 연구 *

정동일 · 이태호
전기 전자 및 자동화 공학부

< 요 약 >

본 연구에서는 여러 개의 HMM(Hidden Markov Model) 원형을 설정하고, 각 원형에 대한 HMM 모형을 발생시켜서, 인식 실험을 통하여 각 모형에 대한 성능을 비교 평가하였다. 사용된 음성 데이터는 15명의 화자로부터 수집한 총 150개의 연속 숫자음이며, 이중 100개는 HMM 모형 발생 작업에 사용된 시험 데이터이고, 나머지 50개는 인식 작업에 사용된 시험 데이터이다. HMM 모형을 발생시키기 위해 13개의 HMM 원형을 설정하였다. 본 연구에서는 두 종류의 HMM 모형 발생 실험을 하였다. 하나는 음소 한 개당 한 개의 모형을 발생시키는 개별 음소 모형 실험이고, 다른 하나는 유사한 특징을 가지는 음소를 묶어서 하나의 모형으로 발생시키는 집단 음소 모형 실험이다. 집단 음소 실험은 개별 음소 모형 실험에서 높은 인식률을 기록한 원형 1, 2, 3, 4, 6을 사용하여 HMM 모형을 발생하였다. 또 개별 음소 실험에서 발생된 HMM 모형의 연결을 바탕으로 단어 인식 실험을 수행하였다. 단어 인식 실험에서 사용된 개별 음소 HMM 모형 또한 개별 음소 모형 실험에서 높은 인식률을 기록한 원형 1, 2, 4, 6에서 발생된 HMM 모형을 사용하였다. 각 실험에서 HMM 모형을 발생시키고 인식률을 산출한 결과 5 상태 좌-우 구조에서 스트림의 수가 2개이고 혼합의 수가 각각 4인 모형이 가장 우수하였다.

A Study on the HMM-Based Phoneme Models for Korean Speech

Dong-Il Chung , Tai-Ho Lee
School of Electrical Engineering and Automation

* 본 연구는 1998년 울산대학교 학술 연구비의 지원으로 이루어졌음.

<Abstract>

In this paper, the Hidden Markov Models(HMMs) which are appropriate for Korean speech are studied. A set of prototypes is prepared and a set of HMMs is generated from these prototypes. These HMMs are assessed by the recognition test of Korean phonemes. The speech data used in this paper are the utterances of the Korean numbering system from 'hana(one)' to 'yul(ten)' which are taken from fifteen speakers. Every speaker pronounces these numbers ten times so that the total number of the sets of utterances is one hundred fifty. One hundred sets among those data are used for generating and training the HMM and the others are used for testing the HMMs. Two different types of HMM phoneme models are studied. One is the individual phoneme model. In this case, a model corresponds to a phoneme. The other is the large category model which represents a group of phonemes. Along with the phoneme recognition, the work level recognition test is performed. The HMMs for a word are constructed by linking the individual HMM phoneme models. The best result is obtained from the model which has the five states, left-right structure with two streams and four mixtures in each stream.

I. 서 론

디지털 기술의 발달로 인간과 기계간의 의사소통(man-machine interface)의 중요성이 크게 대두되고 있다. 음성도 이러한 의사소통에 있어서 가장 기본적인 매개체중의 하나이며 그런만큼 오랫동안 관심을 받아 오고 있다. 그러나, 많은 연구에도 불구하고 음성인식 및 이와 관련된 음성처리 기술이 빠른 진전을 보이지 못하고 있는데, 그 가장 큰 이유는 음성정보가 본래적으로 대단히 애매하고 가변성이 많기 때문이다.[1][2][10] 현재의 음성인식의 연구에는 패턴정합에의한 방법[4], 은닉마르코프모델(이하 HMM)[1][2][6][7][13][14]및 신경회로망[5]으로 처리하는 연구가 주축을 이루고 있다. 패턴 정합(pattern matching)에 의한 방법은 입력음성에서 특징벡터를 추출한 다음, 기준음성들과 비교하여 가장 유사도가 높은 기준음성을 선택하는 방법이다. 주로 동적 계획법을 사용하는데 기준 음성의 구축이 쉽고 인식률이 높은 장점이 있으나 계산량이 많고 연결단어의 인식으로의 확장이 어렵다. 신경회로망은 훈련 신호들을 학습하여 원하는 입출력간의 정합에 의해서 패턴을 분류하는 방식으로 인식시 계산량이 적고 속도가 빠르다는 장점이 있으나 소규모 고립단어 또는 음소인식과 같은 일부분에서만 연구 결과가 발표되고 있다. HMM을 사용한 인식법은 음성의 특질을 통계적으로 처리하고, 이 통계량을 확률 형태의 모델에 반영하여 인식하는 방법이며, 확률 모델을 사용하므로 여러 가변적 요소를 쉽게 반영할 수 있고 음소나 음절단위의 모형을 쉽게 단어 문장등의 단위로 확장할 수 있는 장점이 있다[1][2]. 이 인식법은 기준 음성에 대한 HMM 모형을 만든 후, 이들 모형으로부터 가장 높은 확률을 가지는 모형을 선택하는 방법으로, 인식단위에 대한 적절한 모형만 만들어진다면 상당히 큰 효과를 발휘 할

수 있다.

음소단위의 인식기법(rule-based method)[12][13]은 비록 많은 작업과 시간을 요구하는 방법이긴 하나 불특정화자의 대용량 어휘인식에서는 음소단위의 인식 기법이 상당히 유리하다.

일반적으로 어휘 규모가 크고, 특히 화자 독립의 경우라면 HMM 모형을 만들기 위해서는 상당량의 훈련용 음성자료가 요구되는데 이 음성자료에 대해서는 구체적인 음성정보가 공급 되어야 한다. 예를 들어 음소 모형을 만들려면 각 음성자료를 음소별로 분절한 음소기술(phoneme transcription)이 필요하게 된다. 음소 분절 작업은 대단히 많은 인력과 시간을 필요로 하므로 착수하기가 쉽지 않다.

본 연구에서는 몇 개의 다른 구조의 HMM 모형에 대하여 한국어 음성의 음소 인식 및 이를 바탕으로 한 단어 인식 실험을 수행하여 그 성능을 비교하였다. 본 연구에서 사용한 음성데이터는 15명의 화자가 10번씩 발성한 연속 숫자음이며 이 중 10명에 대한 음성은 학습데이터로 5명에 대한 음성은 시험데이터로 사용하였다. 분절작업은 100개의 학습 데이터를 수 작업으로 분절하였다. 분절된 데이터를 이용하여 개별 음소에 대한 HMM 모형과 집단 음소에 대한 HMM 모형을 발생시켜 실험을 하였다.

HMM은 관측 벡터가 이산적이고 유한한 경우의 이산관측 HMM과 연속적인 경우의 연속관측 HMM의 두 가지로 구분되는데 본 논문에서는 연속 HMM을 사용하였다.

HMM 모형을 만들어내기 위해서는 HMM 원형 (prototype)을 설정해야 한다. 본 연구에서는 6상태, 5상태 좌우구조의 두 가지 HMM 구조를 설정하였고, 세부적으로는 스트림의 수, 스트림내의 혼합의 수, 각 상태간의 천이에 변화를 주는 방식으로 각각 13개의 원형을 만들어 사용하였다.

설정된 원형을 이용하여 초기화작업과 Baum-Welch 재추정 방법을 거쳐 최종 HMM 모형이 된다. 얻어낸 HMM 모형을 이용하여 음소단위 인식(phoneme level recognition)작업을 수행하였으며 개별 음소 모형 실험에서 발생된 HMM모형으로 단어 인식(word level recognition)을 시도하였다. 인식작업의 결과는 수 작업으로 분절된 데이터와 자동 분절된 데이터의 음소 배열 패턴을 비교하여 산출하였으며, 이를 바탕으로 발생된 모형을 평가하였다. II장에서는 HMM에 대한 기본 이론을 다루고, III장에서는 본 연구에서 구현한 시스템 및 실험에 대한 내용을 정리하였다. IV장에서는 실험한 결과 및 분석한 내용을 기술하였으며, V장에는 본 연구에 대한 결론 및 향후 개선 방향을 제시하였다.

II 은닉 마르코프모델(Hidden Markov Model)

2.1 HMM

HMM은 다중 확률 구조를 갖는 프로세스들을 모델링하는 데 매우 적합하며 HMM 파라미터들의 정밀한 계산을 위한 효과적인 알고리듬이 존재한다는 장점때문에 근래에 가장 많이 쓰이는 음성인식 기법으로 Baker와 IBM 연구진에 의해 제안되었고, 그것에 관한 이론은 Baum 등의 연구에 기초를 두고 있다.

HMM은 확률의 상태와 그들간의 천이 확률로 정의되며, 각 천이는 상태선택에 관한 천이 확률, 천이가 이루어졌을 때 유한개의 관측 심볼로부터 각 출력 심볼이 방출되는 조건부 확률에 대한 확률함수라는 2가지 종류의 확률과 관련되어 있다. 즉 HMM은 관측이 불가능한 하나의 프로세스를 관측이 가능한 심볼로 발생시키는 프로세스를 통하여 추정하는 이중확률 프로세스로서 음성과 같이 가변성이 많고 발생과정을 알 수 없는 프로세스들을 모델링하는데 적합하다.

음성 발성에서 음성의 발성구조를 마르코프 체인의 각 상태로 간주하고 성도가 그 상태 중 하나에 존재한다고 가정하면, 각 음성 발화에 대한 성도의 전달특성은 관측이 불가능한 프로세스에 해당하며 발성된 음성신호는 관측 가능한 프로세스가 되므로, 발성된 음성신호를 위와 같은 구조를 가진 HMM에서 나온 것으로 간주할 수 있다. 이 경우 각 상태에서는 그 상태에 해당하는 스펙트럼을 지닌 음성 세그먼트가 발생하게 된다. 이러한 음성 세그먼트는 음성의 발성 분포를 표현하는 파라미터로 이루어져 있으며 음성 주파수의 변동성을 나타낼 수 있다. 그리고, 마르코프 체인의 상태변화는 음성신호 스펙트럼의 시간에 따른 변동성을 나타낼 수 있다.

HMM은 관측벡터가 이산적이냐 또는 연속적이냐에 따라 이산 HMM과 연속 HMM의 두 가지로 구분된다. 이산 HMM은 전처리를 요구하는데 전처리 기법으로는 보통 벡터 양자화를 많이 사용하고 있다[14]. 이 경우 양자화 과정에서 발생하는 오차 발생과 인식과정 중 그 오차가 계속 전달되는 오류 전달 현상이 발생할 수 있다. 연속 HMM은 확률 밀도 함수로 가우시안 분포함수를 사용하므로 확률 계산에 따른 계산량이 방대한 단점이 있다. 본 논문에서는 인식률이 다소 높은 연속 HMM을 사용하였다.

2.2. HMM에 대한 파라미터와 알고리듬

2.2.1 HMM

HMM은 복잡한 다중 확률 분포를 가지는 물리 현상을 모델링하는데 부적절한 마르코프 체인을 보완하기 위해 1960년대말 Baum등이 확장한 모델이다. HMM은 상태 천이 확률, 관측 심볼의 확률분포 및 초기 상태확률의 3가지 확률 파라미터로 정의 되며 다음과 같이 표현될 수 있다.

$$\lambda = (A\ ,\ B\ ,\ \Pi) \tag{2-1}$$

A는 상태 천이 확률 분포를 나타내는 상태천이 행렬이고, 다음과 같다.

$$A = [a_{ij}]\ ,\quad 1 \le i \le N \tag{2-2}$$

a_{ij}는 상태 i에서 상태 j로 천이될 확률이고 N은 상태 수이다.

B는 관측심볼의 발생확률 분포를 나타내며 다음과 같다.

$$B = [\ b_j(k)\ ]\ ,\quad 1 \le j \le N\ ,\quad 1 \le k \le M \tag{2-3}$$

$$b_j(k) = P\ [\ o_t = v_k \mid q_t = j\ ] \tag{2-4}$$

$b_j\ (k)$는 현재상태 q_k에서 관측심볼(o_t)이 v_k가 될 확률이며 M은 관측 심볼의 개수이다. B가 이산 확률 분포일 경우 이산 HMM이 된다.

마지막으로 Π는 초기 상태확률 분포이다.

$$\Pi = [\pi_1, \pi_2, \pi_3, , , \pi_N] \tag{2-5}$$

위의 만들어진 HMM은 전 후향 알고리듬, Viterbi 알고리듬, Baum-Welch 재추정 알고리듬[1][2][3]을 이용하여 실제 문제에 적용되어질 수 있다.

전 후향 알고리듬은 관측열 $O=[O_1, O_2, , , O_T]$와 HMM $\lambda=(\pi A B)$가 주어졌을 때 관측열을 발생시킬 확률 $P(O|\lambda)$를 계산하는 확률계산법이다. Viterbi 알고리듬은 최적의 상태열과 그 상태열을 통한 확률을 구하는데 사용되고, Baum-Welch 재추정 알고리듬은 초기 모델이 주어졌을 때, 학습 데이터를 사용하여 관측심볼의 발생확률을 최대화하기 위해 HMM 파라미터 λ를 반복적으로 학습시키는데 사용한다.

2.2.2 연속 HMM

연속 HMM은 관측 벡터가 연속적인 경우로서 관측 확률 밀도 B가 연속적이다. B를 표현하는 대표적인 모형은 가우시안 함수 형태이며 그림 2-1과 같이 유한개의 단일 혼합 가우시안(Single mixture gaussian)함수 또는 부-상태 가우시안 함수(Sub-state gaussian)의 합으로 표현된다. 여기서 각 가우시안 함수는 서로 다른 혼합계수 c_{ik}를 가지도록 한다.[1][2]. 즉, 가중치를 달리하는 가우시안 분포를 이용하여 실제 관측 심볼의 분포를 최대로 근사화한 것이다. 혼합계수 c_{ik}에 대한 제약은 다음과 같다.

$$\sum_{k=1}^{M} c_{jk} = 1, \quad 1 \le j \le N \tag{2-6}$$

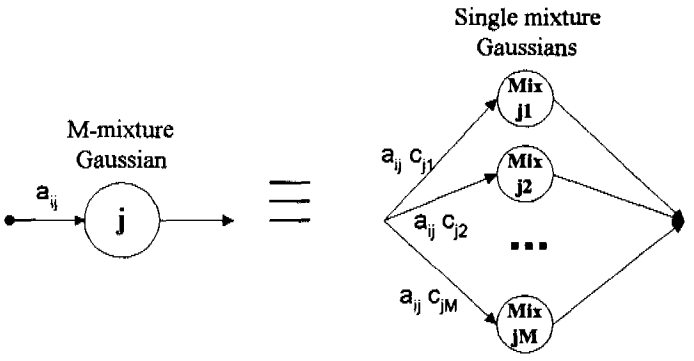


그림 2-1. 다중 혼합요소에 대한 부상태 표현
Fig. 2-1. Sub-state Representation of Multiple Mixture Components

연속 관측 확률밀도 함수의 일반적인 형태는 다음과 같다.

$$b_j(o) = \sum_{k=1}^{M} c_{jk} N(o, \mu_{jk}, U_{jk}), \quad 1 \le j \le N \tag{2-7}$$

N은 가우시안 확률 분포함수를 나타내며 식 2-8과 같다.

$$N(o, \mu_{jk}, U_{jk}) = \frac{1}{\sqrt{(2\pi)^n |U|}} e^{-\frac{1}{2}(o-\mu)' U^{-1}(o-\mu)} \tag{2-8}$$

여기에서, μ_{jk}는 평균 벡터를 U_{jk}는 공분산 행렬을 나타내고, n은 관측 벡터의 차수이다. c_{jk}는 각 가우시안 확률 분포함수에 해당하는 혼합계수이다.

연속 은닉 마르코프 모델에 대한 파라미터 재추정 방식은 이산 마르코프 모델일때와 동일하며, 관측심볼 발생 확률에 대한 재추정은 가우시안 함수의 혼합계수와 평균벡터, 공분산행렬의 재추정 문제가 된다. 각 부-상태 가우시안 함수의 파라미터 재추정식은 다음과 같다.

$$\overline{c_{jk}} = \frac{\sum_{t=1}^{T} \gamma_t(j,k)}{\sum_{t=1}^{T} \sum_{k=1}^{M} \gamma_t(j,k)} \tag{2-9}$$

$$\overline{\mu_{jk}} = \frac{\sum_{t=1}^{T} \gamma_t(j,k) \cdot o_t}{\sum_{t=1}^{T} \gamma_t(j,k)} \tag{2-10}$$

$$\overline{U_{jk}} = \frac{\sum_{t=1}^{T} \gamma_t(j,k) \cdot (o_t - \mu_{jk}) \cdot (o_t - \mu_{jk})'}{\sum_{t=1}^{T} \gamma_t(j,k)} \tag{2-11}$$

$\gamma_t(j,k)$는 시간 t에서 상태 j의 k번째 혼합 성분을 가지는 부-상태에 존재하는 확률(likelihood) 또는 상태 점유(State occupation)로 정의되며, 다음 식2-12와 같다.

$$\gamma_t(j,k) = \left[\frac{\alpha_t(j)\beta_t(j)}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)}\right] \cdot \left[\frac{c_{jk} N(o_t, \mu_{jk}, U_{jk})}{\sum_{m=1}^{M} c_{jm} N(o_t, \mu_{jm}, U_{jm})}\right] \tag{2-12}$$

2.3 HMM의 구조

HMM은 그 구조적 형태와 상태내의 파라미터로 정의된다.

HMM의 구조적 형태는 모델내의 각 상태에 대한 상태천이 확률을 나타내는 상태 천이 행렬 A에 의해 결정된다. HMM의 형태는 자신 및 모든 상태로 천이가 가능한 엘고딕 모델(ergodic model), 시간의 흐름에 따라 상태천이를 일정한 방향으로만 진행하도록 제한한 좌-우 구조 모델(left-right model), 두 개이상의 방향으로 제한한 병렬 좌-우 구조 모델(parallel path left-right model)등이 있다. 음성신호에는 신호 자체가 시간에 따라 연속적으로 변하므로 음성인식에는 좌-우 구조가 적합하다. 본 연구에는 5상태 및, 6상태 좌-

우 구조 모델을 사용했다.

HMM의 내부 상태내의 파라미터는 각 상태의 천이확률, 각 상태에서의 관측 심볼 확률분포 및 각 상태내의 혼합의 수 등으로 정의된다. 그 예로, 그림 2-2에 좌-우 구조 및 각 파라미터로 정의된 3상태 연속 HMM을 나타내었다.

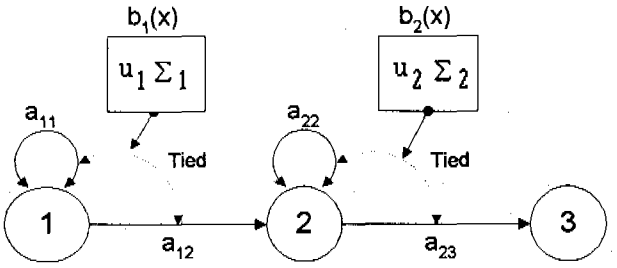


그림 2-2. 3상태 HMM
Fig 2-2. 3 states HMM

Ⅲ. 실험

3.1 HMM을 이용한 음성인식 및 발생된 HMM 모형 평가 시스템

그림 3-1은 본 연구에서 구현한 시스템의 블록도이다.

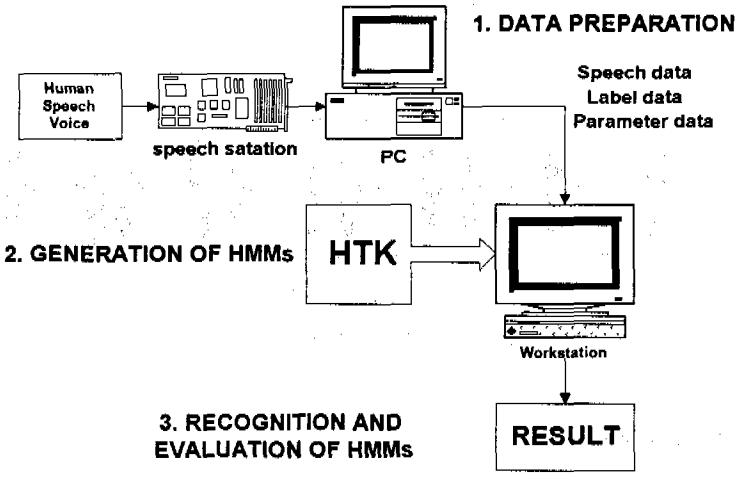


그림 3-1. HMM을 이용한 인식 시스템
Fig 3-1. Recognition System using HMM

본 실험에서는 시스템을 크게, 데이터 준비 작업, HMM 모형 발생작업, 인식 및 HMM 모형 평가 작업으로 나누어 실험하였다.

3.2 데이터 준비 작업

실제 화자가 발성한 음성 데이터를 수집하고, 이 음성신호를 분절한 레이블 데이터를 작성한 다음, 특징벡터를 포함한 파라미터 데이터를 발생하여 실험하였다.

3.2.1 음성 데이터 수집

남성화자 10명, 여성화자 5명이 각각 10번씩 발성한 150개의 음성 데이터를 사용하였다. 화자 독립을 원칙으로 하여, 총 150개중 10명의 화자에 대한 데이터 100개는 HMM 모형에 대한 학습과정에, 나머지 5명의 화자에 대한 데이터 50개는 시험과정에 사용하였다. 음성 데이터 베이스 구축도구로 미국 Sensimetrics 회사 제품인 Speech-Station 3.0을 사용하였고, 데이터 획득 과정의 파라미터는 표3-1과 같다.

Sampling rate	16000 Hz, 16bit
Recoding time	7.8sec
입력 장치	Static microphone
Gain	1

표 3-1. 데이터 획득 변수
Table 3-1 Parameters for Data Acquisition

화자의 발성음은 "하나"부터 "열"까지의 연속 숫자음이다. 그림 3-2는 발성된 음성 파형을 나타내고 있다.

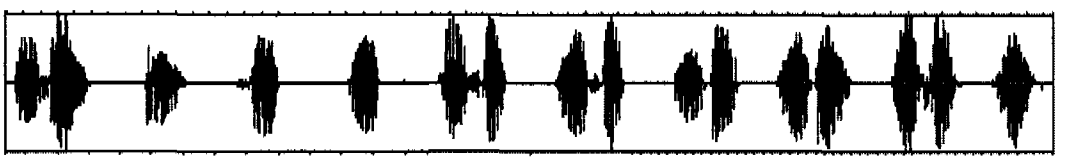


그림 3-2. 음성 파형
Fig 3-2. Speech waveform

3.2.2 데이터의 레이블링

수집된 음성데이터에 대해 각 음소별로의 시간적 순서와 각 음소간의 경계를 기술한 데이터를 작성하는 작업이다.

이 작업의 과정은 분절 작업과 각 구간에 음소의 명칭을 기록하는 작업으로 구성된다.

초기 레이블작업은 스펙트로그램상의 스펙트럼의 형태와 시간파형에서 나타나는 경계를 유심히 관찰하여 분절한 다음, 각 구간에서 나오는 소리를 직접 듣고 수정을 되풀이하면서 수행하였다. 분절된 각 구간에 대한 음소의 명칭을 붙이는데 이 작업을 레이블링이라 한다. 표3-2에 "하나"부터 "열"까지의 연속 숫자 음성에 대해 본 실험에서 표시한 레이블을 나타내었다.

하나	둘	셋	넷	다섯	여섯	일곱	여덟	아홉	열
[ha:na]	[tul]	[secld]	[necld]	[tas^cld]	[j^s^cld]	[ilgoclb]	[j^d^l]	[ahoclb]	[j^l]

표3-2. 숫자음에 대한 레이블

Table 3-2. Labels for spoken numbers

소리가 없는 구간인 묵음(silence)은 [x]로 표시하였고, [a:]는 "ㅏ"음의 장모음에 해당한다. 종성에서 [ㄷ]로 끝나면 [cld]로 [ㅂ]으로 끝나면 [clb]로 표시하였다. "ㅕ"발음인 경우는 "반모음 ㅣ"[j] + 모음 "ㅓ"[^]로 구분하여 표시하였고, 초성 'ㄷ'인 경우 "다섯"에서와 같이 무성음이 강한 부분은 [t]로, "여덟"과 같이 유성음이 강한 부분은 [d]로 표시 하였다.

본 실험에서는 개개의 음소별로 하나씩 HMM 모형을 만들어 훈련시키는 개별 음소 모형 실험과 비슷한 특징을 가지는 음소끼리 묶어서 하나의 모형을 발생시키는 집단 음소 모형 실험의 두 가지를 수행하였다.

3.2.2.1 개별 음소 모형

개별 음소 모형 실험에서는 [a:]와 [a]를 [A]로, [d]와 [t]를 [D]로 묶었으며 나머지는 개별 음소 당 하나의 모형을 가지도록 분류하였다.

16개로 분류된 각 음소의 명칭은 표 3-3에 나타내었다.

h	A	D	s	n	e	j	^	i	l	g	o	cld	clb	u	x
ㅎ	ㅏ	ㄷ	ㅅ	ㄴ	ㅔ	반모음 ㅣ	ㅓ	ㅣ	ㄹ(종성)	ㄱ	ㅗ	ㄷ(종성)	ㅂ(종성)	ㅜ	묵음

표 3-3. 분류된 각 음소의 명칭

Table 3-3. Each name of segmented phonemes

그림3-3과 그림 3-4는 "하나"음에 대한 분절 데이터와 개별 음소 모형으로 분절된 음성 파형의 형태와 모습을 나타내고 있다.

```
0 1220000 h
1220000 3059999 A
3050000 3619999 n
3619999 5960000 A
5960000 9959999 x
```

그림 3-3. "하나"음에 대한 레이블 데이터
Fig 3-3 Labled data for word 'hana'

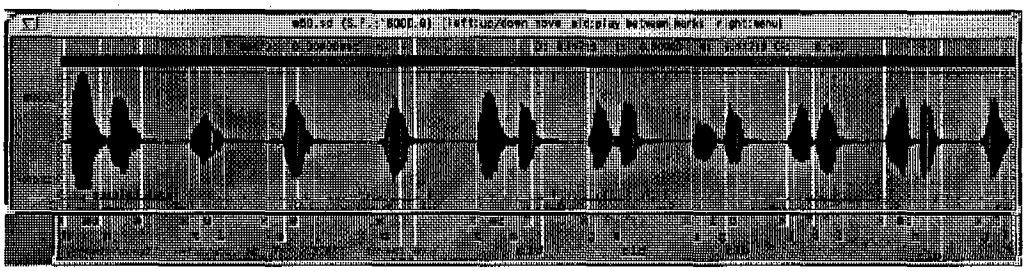


그림 3-4. 개별 음소 모형으로 레이블된 음성데이터
Fig 3-4. Speech waveform labeled by each phoneme model

3.2.2.2 집단 음소 모형

집단 음소 모형 실험에서는 모음, 유성자음, 무성자음, 정지자음, 묵음 등 6가지 형태의 대표음소를 사용하여 분류하였다. "반모음 ㅣ"를 제외한 모든 모음은 V로 표기 하였고, 자음중 비음과 유음 그리고 "반모음 ㅣ"[j]를 유성자음으로 분류하여 VC로 표시하였다. 나머지 자음중 [s], [h], [t]는 무성자음으로 분류하여 SC로 표기했고, 종성이면서 정지음을 가지는 [cld] [clb]는 정지자음으로 분류하고 CL로 표기하였다. 집단 음소 모형으로 분류된 형태를 표3-4에 나타내었다.

대표음소의 명칭	종류	포함된 음소
V	모음	a a: i u e o ^
SC	무성자음	s h t
VC	유성자음	n d j l g
CL	정지자음	cld clb
x	묵음	x

표. 3-4 집단 음소 모형의 명칭
Table. 3-4 The list of phoneme group models

집단 음소 분류중 비음 [n]의 경우를 따로 분류하여 인식시키는 경우도 있다.[10][15]
그림3-5와 그림 3-6은 "하나"음에 대한 분절 데이터와 집단 음소 모형으로 분류된 음성 파형의 형태을 나타내고 있다.

```
0 1220000 SC
1220000 3059999 V
3050000 3619999 VC
3619999 5960000 V
5960000 9959999 x
```

그림 3-5. "하나"음에 대한 집단 음소 레이블 데이터
Fig 3-5. Labled data for 'hana' by large category

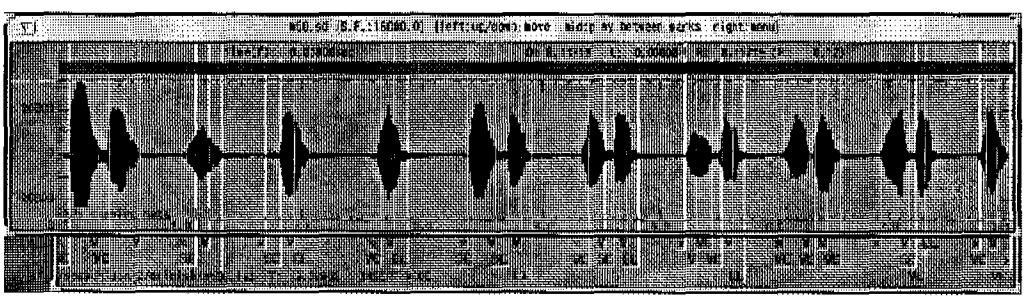


그림 3-6. 집단 음소 모형으로 레이블된 음성파형
Fig 3-6. Speech waveform labeled by large category

3.2.3 파라미터 데이터 발생

음성 신호에 대해 특징 벡터를 추출하는 것은 음성신호에 대한 전처리 과정중의 하나이다.

먼저 음성신호 주파수 스펙트럼의 6dB/oct 감쇄 특성 때문에 줄어드는 고주파 성분을 보상하기 위한 전처리(pre-emphasis) 과정이 필요하다. 본 실험에서는 전처리 계수를 0.97로하여 음성신호를 다음의 1차 디지털 필터를 통과 시켰다.

$$H(z) = 1 - 0.97 \cdot Z^{-1} \qquad (3\text{-}1)$$

음성신호에 대한 특징 벡터는 신호에 윈도우를 씌운 다음 윈도우를 프레임 주기만큼 이동 시켜 프레임단위로 추출된다. 본 실험에서는 윈도우 크기를 25.6 m-secs로, 프레임 주기를 10 m-secs로 하여 15.6 m-secs의 윈도우 겹침을 일어나게 하였고, 윈도우 경계면의 누설오차(leakage error)를 없애기 위해 Hamming 윈도우를 사용하였다. 프레임 단위로 분할된 음성 데이터에 정규화된 대수 에너지(nomalized log energy)를 부가하여 에너지를 정규화 한 다음, 24채널 mel-필터 뱅크 해석을 통해 최종 특징 벡터인 mel-주파수 켑스트럼 계수(MFCC)를 구하였다. 구한 특징 벡터의 형태를 정규화된 대수 에너지 계수가 부가되었으므로 MFCC_E라 정의하였다.

mel-필터 뱅크 해석을 위해 사용된 mel 척도는 다음과 같다.

$$Mel(f) = 2595 \log_{10}(1 + \frac{f}{700})$$

그림 3-7은 mel-척도 필터 뱅크를 나타내고 있다.

특징 벡터인 MFCC $\{c_i\}$는 mel-척도 필터 뱅크를 통과한 값의 대수치(m_j)를 DCT 변환을 이용, 다음과 같이 구하였다.

$$C = \{c_i\}, \quad c_i = \sum_{j=1}^{P} m_j \cos\left(\frac{\pi i}{P}(j - 0.5)\right), \quad 1 \le i \le N \qquad (3\text{-}3)$$

P는 필터 뱅크의 채널 수를 그리고, N은 벡터 크기를 나타내며, 본 실험에서는 P를 24로 N을 12로 설정하였다. 그림 3-8은 특징 파라미터 추출과정을 나타내고 있다.

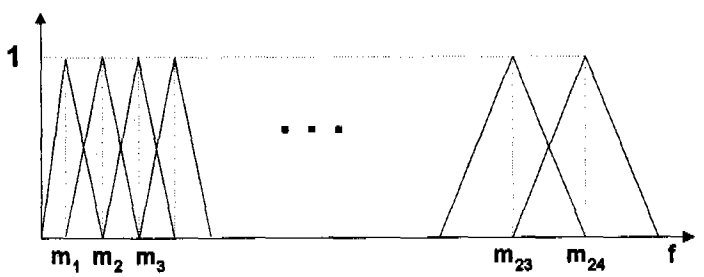


그림 3-7. mel-척도 필터 뱅크
Fig 3-7. mel-scale filter bank

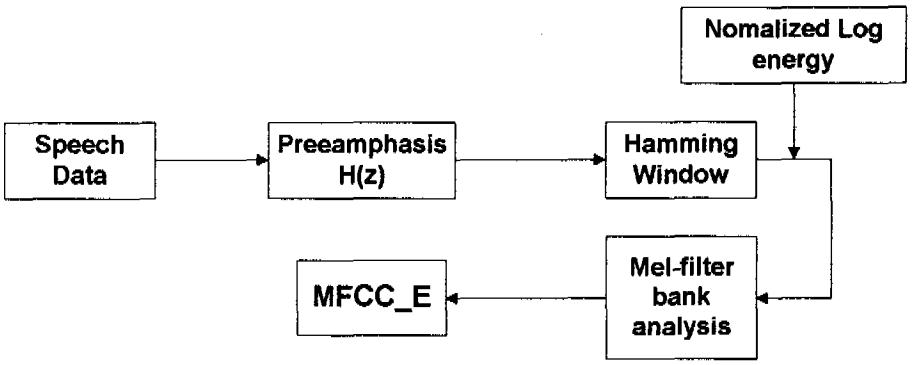


그림 3-8. 특징 파라미터 추출과정
Fig 3-8. Processing for characteristics parameters extraction

3.3 HMM 모형의 발생

3.3.1 HMM을 위한 도구

HMM 모형과 인식 도구로는 미국 Entropic사와 영국 Cambridge 대학의 제품인 HTK 소프트웨어를 사용하였다. 이 도구는 모형의 원형의 발생으로부터 이의 훈련, 그리고 훈련된 모형을 사용한 인식 시험에 이르기까지 각 단계를 망라하고 있다[16]. 본 실험에서의 모형발생 작업 흐름을 그림 3-9에 나타내었다.

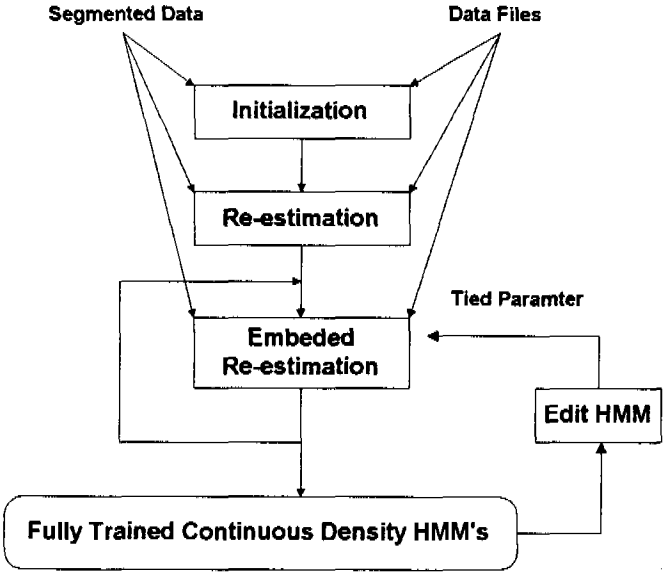


그림 3-9. HMM 모형 발생작업
Fig 3-9. HMM Generation process

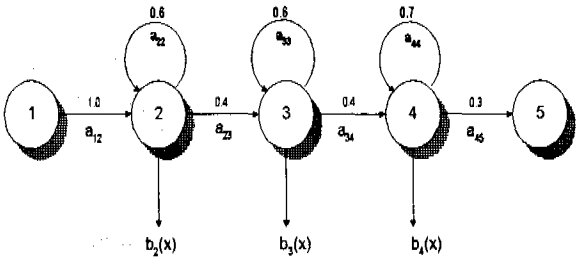
HMM 모형 발생 작업은 HMM 원형 설정 및 초기화, 재추정 및 부가 재추정의 세 가지 학습 단계로 나누어 수행하였다.

3.3.2 HMM 원형 설정

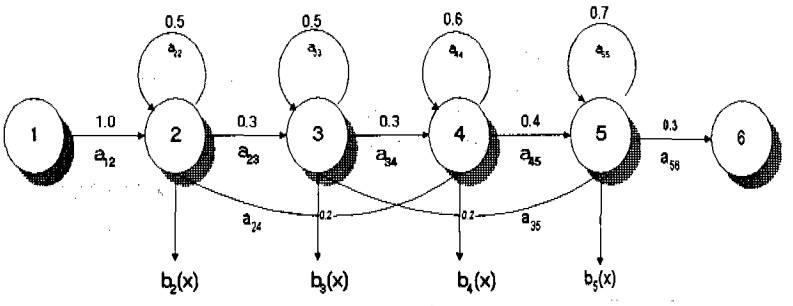
HMM 원형은 만들어낼 HMM의 형태를 결정한다. 본 실험에서 설정한 HMM 형태는 음성인식에서 가장 많이 사용되는 좌-우 구조이며, 상태수는 5개와 6개로 선정하였고, 6 상태 HMM은 대표 음소 모형 실험에서만 사용하였다. 5개 상태의 HMM은 처음과 마지막의 상태의 자체천이가 없으며, 나머지 상태에 대해선 인접 전향천이와 자체 천이만을 가지도록 설정하였다. 6 상태 HMM은 5 상태 HMM과 유사하나, 2와 3 상태에서 각각 4와 5로의 건너뛰는 천이가 발생하도록 설정하였다.

처음과 끝의 천이를 제한한 이유는 차후의 독립적으로 교육된 HMM의 공통 파라미터 연결을 위한 교육과정에서 음소별 모형의 연결 상태로 이용하기 위해서이다.

그림 3-10은 본 실험에서 설정한 HMM 원형 구조와 상태 천이 행렬 형태이다.



a) 5-states HMM



b) 6-states HMM

$$A_5 = \begin{bmatrix} 0 & a_{12} & 0 & 0 & 0 \\ 0 & a_{22} & a_{23} & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} & 0 \\ 0 & 0 & 0 & a_{44} & a_{45} \end{bmatrix} \qquad A_6 = \begin{bmatrix} 0 & a_{12} & 0 & 0 & 0 & 0 \\ 0 & a_{22} & a_{23} & a_{24} & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} & a_{35} & 0 \\ 0 & 0 & 0 & a_{44} & a_{45} & 0 \\ 0 & 0 & 0 & 0 & a_{55} & a_{56} \end{bmatrix}$$

$$A_n = \{ a_{ij} \} \quad , \quad \sum_{j=1}^{n} a_{ij} = 1$$

c) 상태 천이행렬

그림 3-10.. HMM 원형
Fig 3-10. The Prototype of HMM

HMM 원형에는 또한 관측 확률 밀도 함수의 구조와 공분산 행렬의 형태를 설정해야 한다. 관측 확률 밀도 함수는 식 2-7과 같은 혼합 가우시안을 사용하였다. 이것은 다시 관측 대상의 수에 따라 각각 하나씩 정의가 되는데 이것을 스트림이라 한다. 예를 들어 스트림이 2이면 관측 대상이 HMM 출력과 그것의 변화량이 되고 이것에 대한 각각의 가우시안 확률밀도가 정의된다. 본 실험에서는 스트림의 수, 각 스트림 내의 혼합의 수, 그리고 공분산 행렬의 형태를 달리 하면서 총 13개의 HMM 원형을 설정하였다. 표3-5에 설정한 13종의 원형을 나타내었다.

표 3-5. HMM 원형
Table 3-5. The prototypes of HMMs

	1	2	3	4	5	6	7	8	9	10	11	12	13
스트림의 수	2	2	2	2	2	2	3	2	1	2	2	3	3
혼합의 수	2,2	2,3	3,2	3,3,	1,1	4,4	2,2,2	4,3	2	3,4	5,3	2,2,3	2,3,4
공분산 행렬	F	C	F	F	F	F	C	F	C	F	F	F	F

* F : Full covariance matrix , C : Diagonal covariance matrix
* 파라미터 형태 : MFCC_E

개별 음소 모형 실험에서는 5 상태 HMM 구조로만 실험하였으며 13개의 원형 모두를 적용하였다. 집단 음소 모형 실험에서는 5 상태와 6상태 HMM 구조에 대해 실험하였으며 개별 음소 모형 실험에서 좋은 인식 결과를 얻어낸 원형 1, 2, 3, 4, 6을 적용하였다.

3.3.3 HMM의 학습

초기화 작업을 통해, 획득한 데이터들과 HMM 원형을 이용하여 각 음소별로 모형을 초기화된다. 사용되는 알고리듬은 균일 분절(uniformly segmentation)과 Viterbi 정렬(alignment) 및 수정된 k-mean clustering이고, 초기화 과정은 다음과 같다. 먼저 분절 데이터를 읽은 다음, 음성 데이터에서 모형으로 만들어낼 각 음소 부분을 모두 잘라내어 그 데이터를 읽어들인 다음, 음소별로 분절된 각 데이터에 대해 수정된 k-mean clustering 방

법을 사용하여 초기 파라미터 값들을 반복적으로 계산해 낸다. 첫 번째 사이클에서는 훈련될 데이터들이 균일하게 분절(segment)되고, 각 모형의 상태가 해당 데이터 분절들과 정합된 다음 해당 모형에 대한 평균과 분산들이 추정된다. 두 번째 이하 사이클에서는 균일 분절형태가 Viterbi 정렬형태로 대치된다. 초기화는 전 단계의 추정치와 비교해서 오차가 0일 때까지 수행한다.

재추정 단계는 초기화 단계와 거의 유사하고, 추정 학습 방법이 수정된 k-mean clustering 방법대신에 Baum-Welch 재추정 방법으로 대치되어 초기화된 각 음소별 모델을 개별적으로 갱신한다.

부가 재추정 작업은 재추정 작업에 의해 독립적으로 갱신된 모든 음소별 모형을 단일 Baum-Welch 재추정 방법으로 동시에 재추정 한다. 각 훈련될 발성에 대응하는 음소 모형이 상호 연결된 다음, 순차적으로 각 HMM 모형에 대한 상태 점유, 평균, 분산등의 통계치들이 전후향 알고리듬에 의해 축적된다. 축적된 통계치들은 HMM 파라미터를 재추정하는데 사용된다. 마지막 단계를 거친 HMM 모형이 인식 및 평가에 사용될 HMM 모형이다.

3.4. 인식 및 모형 평가

3.4.1 음소 인식

그림 3-11은 인식을 위한 계통도이다. 음소 단위의 인식은 개별 음소 모형 실험과 대표 음소 모형 실험에서 각각 시도되었다. 인식작업의 결과는 수 작업으로 레이블된 데이터와 HMM모형에 의해 자동 분절된 데이터를 1대1로 비교하여 산출하였다.

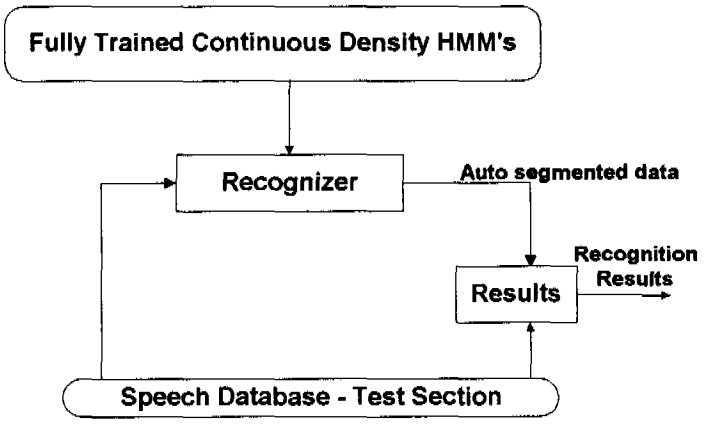


그림 3-11. 인식을 위한 계통도
Fig 3-11. Processing for Recognition

음소 단위 인식은 HMM 모형과 Viterbi 복호 알고리듬, 그리고 파라미터 파일을 이용하여 음성 데이터를 각 구간 별로 가장 확률이 높은 HMM 모형을 삽입하여 분절함으로써 이루어진다. 자동으로 분절된 데이터와 실제 수 작업으로 분절된 데이터를 비교하여 인식률을 산출하였다. 이러한 작업은 두 단계를 거치는데, 첫 번째 단계에서는 HMM 모형 발

생 과정에 쓰인 학습 데이터를 인식한다. 이 작업은 보다 정확한 HMM모형을 만들어 내기 위해 수행하였다. 학습데이터에 대한 인식결과 또는 자동 분절된 결과를 이용하여 수작업으로 분절한 레이블 정보를 갱신하고, 다시 그 레일블링 정보를 이용하여 새로운 HMM 모형을 만들어 내었다. 두 번째 단계는 시험 데이터에 대한 인식 작업으로 최종 발생된 모형을 이용하여 수행하였다.

3.4.2 단어 인식

단어 인식은 개별 음소 모형 실험에서 발생된 음소를 이용하여 수행되었다.

단어 인식에는 크게 2개의 서로 다른 접근법이 있다. 하나는 단어 전체를 하나의 HMM 기본 모형으로 설정하는 방법이며, 또 다른 하나는 음소 등 단어 보다 하위 단위에 대한 HMM을 구축하고 이들의 연결로서 단어를 파악하는 2단계 HMM 방법이다. 전자의 방법이 더 나은 인식률을 나타낼 것으로 예상되지만 단어수가 증가함에 따라 계산량이 크게 높아질 수가 있으며 새로운 단어의 추가 시 전체를 다시 교육해야 되는 등의 문제가 있다. 후자의 경우는 음소 모형의 조합에 있어서 DTW(dynamic time wrapping)를 적용하여 인식률을 높일 수가 있으므로 융통성 면에서 다소 유리하다.

본 연구에서는 후자를 택하였으며 Viterbi 알고리듬으로 2중모형을 인식하도록 하고 있다. 이 때 사용되는 단어 모형은 그림 3-12와 같이 음소 단위의 HMM 모형을 연결을 바탕으로한 네트워크로 정의된다.

모형의 평가 작업은 위의 인식작업의 결과로 나타난 인식률을 이용하여 수행되었다.

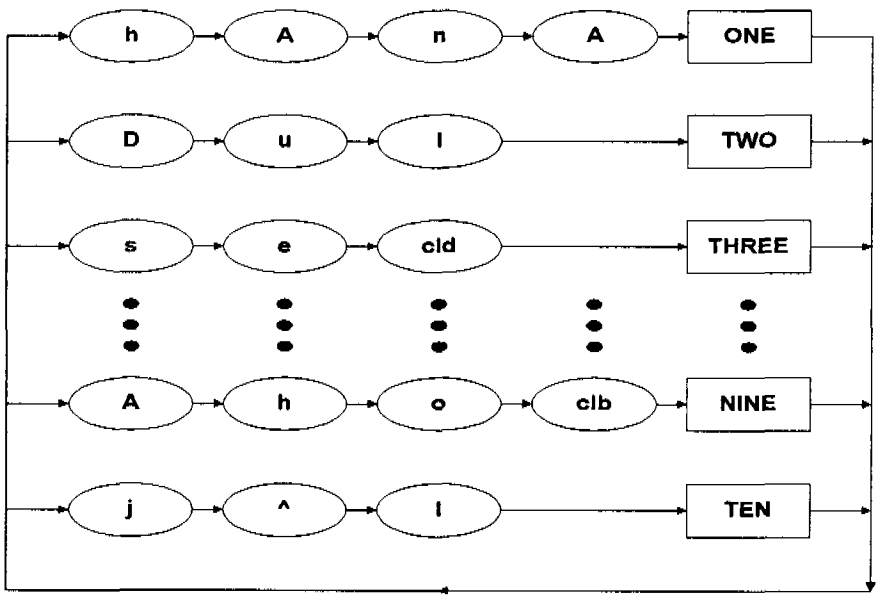


그림 3-12. HMM 연결 network 구조
Fig 3-12. Structure of HMM linking network

Ⅳ. 결 과

4.1 개별 음소 모형 실험

개별 음소 모형 실험에서 산출된 모형을 이용하여 음소 단위 인식과, 단어인식을 시도하였고 결과는 다음과 같다.

개별 음소 모형 실험에서 학습된 각 HMM 모형의 수는 총 5100개이며, 각각의 개별 음소수는 다음과 같다.

[A]	: 4×100 =400개	[u]	: 1×100 =100개	[e]	: 2×100 = 200개
[o]	: 2×100 =200개	[i]	: 1×100 =100개	[h]	: 2×100 = 200개
[j]	: 3×100 =300개	[l]	: 4×400 =100개	[^]	: 6×100 = 600개
[n]	: 2×100 =200개	[g]	: 1×100 =100개	[D]	: 3×100 = 300개
[cld]	: 4×100 =400개	[clb]	: 2×100 =200개	[s]	: 3×100 = 300개
[x]	: 1100개				

표4-1은 최종 학습된 HMM 모형을 발생시켰을 때 학습데이터에 대한 인식률을 나타내고 있다.

표 4-1. 학습데이터에 대한 HMM 모형의 인식률(단위[%])
Table 4-1. The recognition rate of the training data [%]

원형 번호	1	2	3	4	5	6	7	8	9	10	11	12	13
Correct	95.21	95.78	95,64	96.38	86.36	97.36	fail	95.28	90.50	fail	95.83	fail	fail
Accuracy	90.09	91.26	91.05	92.73	82.73	94.12	fail	91.01	84.92	fail	91.47	fail	fail

"fail"는 HMM 부가 재추정 과정에서 확률 파라미터 값 산출 시 오류로 인하여 HMM 모형자체가 만들어지지 않은 경우이다.

표4-2는 최종 학습된 HMM 모형을 가지고 학습에 사용되지 않은 50개의 시험데이터에 대한 인식 실험을 한 결과이다.

표 4-2. 시험 데이터에 대한 인식률 (단위[%])
Table 4-2. The recognition rate of the test data [%]

원형 번호	1	2	3	4	5	6	7	8	9	10	11	12	13
Correct	81.54	83.19	83.01	85.03	76.25	87.53	fail	82.97	78.69	fail	83.71	fail	fail
Accuracy	75.46	76.46	76.12	79.92	70.07	81.06	fail	76.00	72.38	fail	77.04	fail	fail

인식률을 산출해 본 결과, 원형 6에 대한 HMM 모형이 가장 좋게 나왔다. 위에서 산출

된 인식률은 자동 분절된 음성파형 전체에 대한 인식률이다. HMM 모형에 대한 인식 평가는 각 모형별로의 오인식률을 산출해 내는 것이 바람직하다.

모형별로 각 원형에 대한 오인식 개수를 조사하여 표4-3에 나타내었다.

모형 / 원형	h	A	n	D	u	s	e	cld	^	j	i	l	g	o	clb	x
1	8.50	9.00	7.00	7.33	8.00	9.00	5.50	9.50	6.50	9.33	13.00	10.50	11.00	8.00	21.0	10.36
2	7.00	9.25	12.00	8.67	13.00	6.33	9.00	5.50	12.67	11.67	17.00	9.75	10.00	9.00	14.0	12.00
3	8.00	8.75	9.00	8.00	17.00	7.67	10.00	4.50	10.5	10.67	11.0	9.25	14.00	11.00	20.0	9.82
4	6.50	9.25	8.50	6.00	9.00	9.00	10.50	4.75	8.67	11.00	14.00	10.00	19.00	6.00	16.0	10.82
5	14.00	9.75	13.50	10.00	21.00	10.03	11.50	7.00	11.33	8.33	14.00	10.75	21.00	13.50	27.0	14.00
6	5.50	4.00	9.00	5.00	4.00	7.00	9.00	4.00	8.00	9.00	8.00	6.50	14.00	6.00	18.0	7.54
8	6.50	8.50	10.50	12.00	18.00	5.30	13.50	13.00	9.50	10.33	9.00	11.25	17.00	7.00	21.0	11.55
9	11.50	7.75	9.50	9.00	27.00	7.00	10.50	4.75	8.5	15.67	16.00	10.25	22.00	13.00	22.0	11.18
10	8.00	8.75	8.50	11.00	16.00	8.67	9.00	4.25	9.33	8.66	15.00	10.25	11.00	10.50	19.0	8.91

표 4-3. 각 모형별 오인식률 (단위[%])

Table 4-3. The misrecognition rate of each Phoneme [%]

위의 산출결과는 특정모형의 자리에 다른 모형이 삽입된 것의 수를 조사하여 계산하였다. 오인식되는 경우를 살펴보면, 첫 시작부분에 설정되어있는 묵음구간이 무시되고 무성자음 [h]으로 인식되는 경우가 많았고, 무성자음 [D, s, h]들간의 상호 혼동이 많이 발생하였다. 또 "일곱" 부분의 "일ㄱ"을 한개의 모형으로 삽입되는 경우가 많이 발생하였다. 그리고 특정 음소구간이 나뉘어져서 동일한 음소가 여러 번 삽입되어있는 경우는 오인식 되는 경우로 보지 않았다.

전체 음성에 대한 인식률은 각 개별 모형에 대한 인식률보다 낮음을 알 수 있다. 자동 분절된 인식 데이터의 경우 한 개의 음소가 들어갈 구간에 동일 음소가 여러번 삽입되는 빈도가 높아서 삽입 음소의 수가 매우 많다. 이 때 전체 음성에 대한 인식률 계산은 이러한 삽입음소의 수를 기준으로 산출하므로 실제 개별 음소에 대한 인식률 보다 낮게 나타난다.

가장 인식률이 높은 모형은 원형6에 대한 HMM모형이었다. 개별 모형별로 살펴보면 [n][c], [^]에 대해서는 원형1, [g][clb]에 대해서는 원형2, [o]에 대해서는 원형 4와 6, [s]에 대해서는 원형8, 그 외 모형에 대해서는 원형 6에 대한 모형이 가장 높은 인식률을 나타내었다.

그림 4-1은 시험 음성 데이터에 대해 수작업으로 분절된 모습과 자동 분절된 모습을 나타내고 있다.

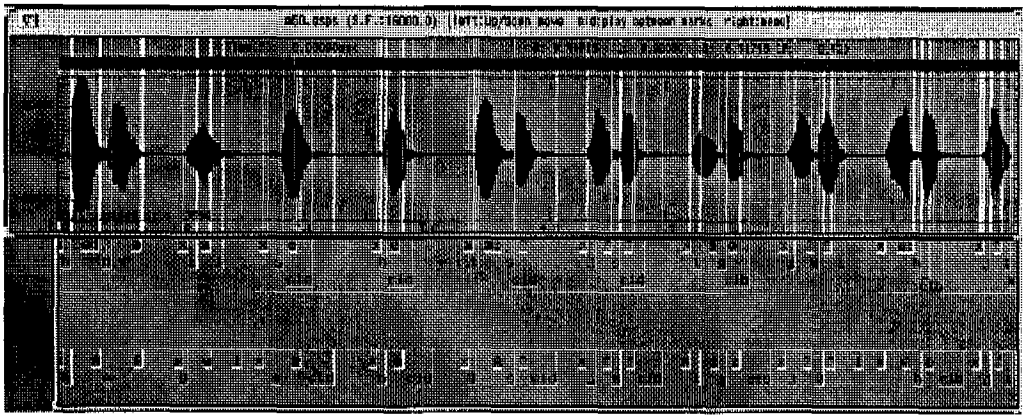


그림 4-1. 수작업 및 자동분절에 의해 레이블된 음성파형
Fig 4-1. Speech waveform segmented by hands and auto-segmentation

4.2 집단 음소 모형 실험

집단 음소 모형 실험에서 학습될 각 HMM 모형에 대한 음소의 수는 다음과 같다.

V	: 16	×	100	=	1600개
SC	: 7	×	100	=	700개
VC	: 11	×	100	=	1100개
CL	: 6	×	100	=	600개
x	:				1100개

표4-4는 최종 학습된 HMM 모형을 발생시켰을 때, HMM 상태 구조에 따른 학습데이터에 대한 인식률을 나타내고 있다. 집단 음소 모형 실험에서는 개별 음소 모형에서 인식률이 높은 1, 2, 3, 4, 5 원형에 대하여 실험하였고, 6상태 HMM에도 같은 원형으로 실험하였다.

a) 5 상태 HMM 모형

원형 번호	1	2	3	4	6
Correct	92.89	93.52	93.45	94.29	95.23
Accuracy	89.92	90.45	90.20	90.36	92.12

b) 6 상태 HMM 모형

원형 번호	1	2	3	4	6
Correct	90.51	91.35	91.21	91.34	91.47
Accuracy	86.78	86.86	86.44	87.09	87.48

표 4-4. 학습데이터에 대한 인식률(단위[%])
Table 4-4. The recognition rate of the training data [%]

표4-5는 최종 학습된 HMM 모형을 가지고 학습에 사용되지 않은 50개의 시험데이터에 대한 인식 실험을 한 결과이다.

a) 5 상태 HMM 모형

원형 번호	1	2	3	4	6
Correct	81.31	8152	81.23	82.05	83.65
Accuracy	71.66	71.57	71.05	72.42	75.88

b) 6 상태 HMM 모형

원형 번호	1	2	3	4	6
Correct	78.69	79.23	79.12	79.49	80.82
Accuracy	69.71	70.37	70.83	71.67	72.52

표 4-5. 시험 데이터에 대한 HMM 모형의 인식률(단위[%])
Table 4-5. The recognition rate of the test data [%]

표4-6에는 각 HMM 모형별로의 오인식률을 나타내었다.

a) 5-states

원형 \ 모형	V	VC	SC	CL	X
1	12.56	16.00	19.14	13.00	11.46
2	11.81	16.46	18.14	13.67	11.36
3	10.75	16.00	18.71	12.57	10.64
4	11.93	16.73	20.14	10.14	11.18
6	9.90	8.18	15.14	7.57	8.27

b) 6-states

원형 \ 모형	V	VC	SC	CL	X
1	11.50	18.46	21.14	15.33	13.91
2	12.00	16.91	23.57	16.00	12.82
3	10.68	17.46	23.28	14.50	15.27
4	11.44	18.82	20.57	13.17	13.10
6	10.63	16.10	21.57	11.33	11.00

표 4-6. 각 모형별 오인식률(단위[%])
Table 4-6. The misrecognition rate of each phoneme[%]

오인식률 산출 방법은 개별 음소 모형 실험에서와 동일하다. 실제 자동분절은 설정된 분절단위로 이루어지는데 연속하는 두 개의 연속 모음일 경우 하나로 인식하는 경우가 많았고, 하나의 묵음구간을 여러 개의 묵음 구간으로 분절되어진 것도 많았다. 그리고 유성자음과 모음은 서로간의 경계가 그리 뚜렷하지 않기 때문에 혼동이 생기는 경우가 많았다. 특히 유성자음을 모음으로 인식하는 경우가 많았다. 각 모형과 원형 상태별로 분석해 본 결과 상태수가 5개인경우에서 인식률이 높았고, 그 중 원형 6 [스트림의 수 2, 혼합의 수 4, 4] 에 대한 모형이 가장 우수한 것으로 나타났다.

그림 4-2는 자동 분절된 시험 데이터를 나타내고 있다.

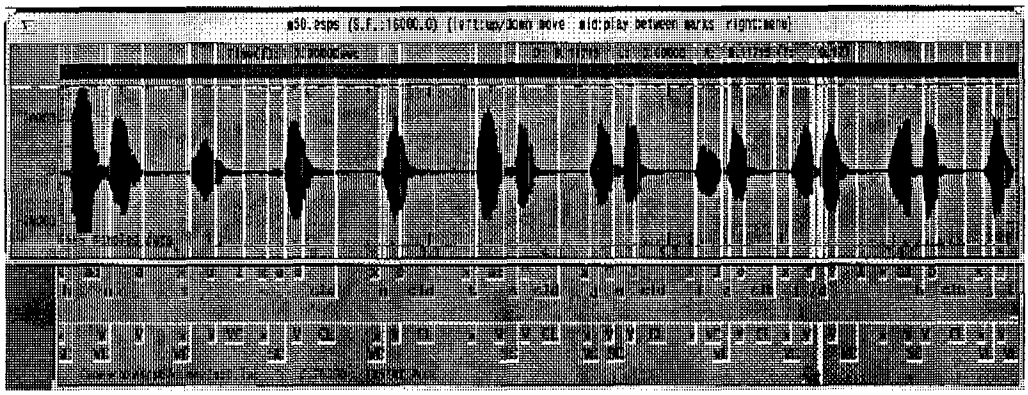


그림 4-2. 수작업 및 자동분절에 의해 레이블된 음성파형
Fig 4-2. Speech waveform segmented by hands and auto-segmentation

4.3 단어 인식 결과

단어 인식작업은 원형 1, 2, 3, 4, 6을 이용하여 수행하였다.
표 4-7은 학습 데이터와 시험 데이터에 대한 단어 인식률을 나타내고 있다.

원형번호 / 데이터	1	2	3	4	6
학습 데이터	98.23	98.75	98.23	98.36	98.67
시험 데이터	90.64	91.50	91.77	91.48	91.19

표 4-7. 학습데이터와 시험 데이터에 대한 단어 인식률(단위[%])
Table 4-7. The word recognition rate of the training and test data [%]

단어 인식작업에서 인식률은 음소 단위에서의 인식률보다 훨씬 좋게 나왔다. 이는 음소 단위인식작업에서 발생된 HMM모형이 쉽게 단어 인식으로 확장될 수 있음을 나타낸다. 단어 인식 과정에서 각 단어 위치는 음소 단위 인식에서 그 단어를 구성하는 음소들의 위치와 일치하였고 음소 단위 인식에서 잘못 인식된 음소들의 상당수가 단어 인식에서는 제대로 인식 됨을 알 수 있었다. 이러한 현상은 음소 단위 인식은 인식될 구간에서는 모든 HMM모형이 자격을 가지고 있으나, 단어 인식 과정은 단어 네트웍구성을 바탕으로 하기 때문에 네트웍에 정의되지 않은 위치에 나타나는 음소는 선택되지 않게 되어 발생한 것 같다. 그러나 단어 인식 과정에서 오인식 된 부분은 그 단어 자체가 없어져 버린 경우가 가장 많았다. 이는 음소단위 HMM 만이 훈련되었고 단어 수준에서 추가적인 교육이 없기

때문에 음소의 변형 및 탈락 등에 대한 처리 능력이 없는데 기인한 것 같다.

그림 4-3은 시험 음성데이터의 단어 인식 결과를 나타내고 있다. 여기서 파형 아래의 레이블 구역에 음소와 병행하여 단어 레이블이 기록된다.

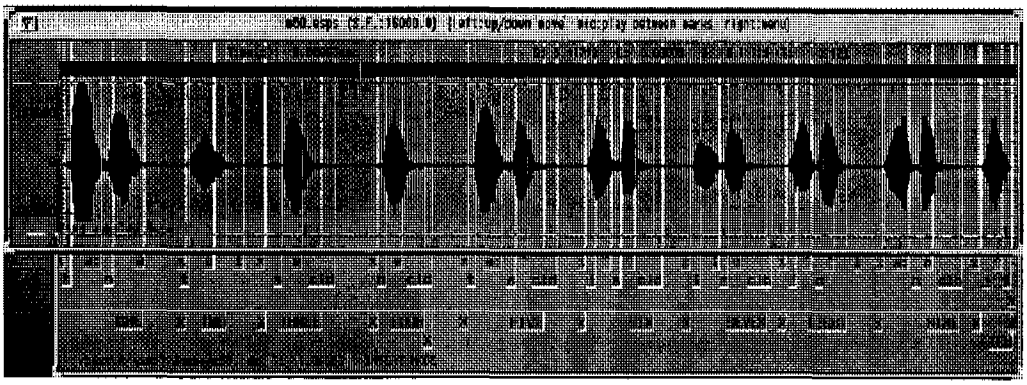


그림 4-3. 시험 데이터의 단어 인식 결과의 예
Fig 4-3. An example of word level recognition of the test data

V. 결론

본 연구에서는 HMM을 이용하여 한국어 음소에 대한 여러 가지 HMM 모형을 발생시키고 이들을 비교 평가하였다.

10명의 화자로부터 얻은 데이터를 이용 HMM 모형을 발생 시켰고, 다른 5명의 화자에 대한 인식 실험을 토대로 각 HMM 모형에 대한 평가 작업을 수행하였다. 본 연구에 나타난 시스템은 크게 "데이터 준비 작업", "HMM 모형 발생", " 인식 및 모형 평가"로 나뉜다.

개별 음소 모형 실험에서는 스트림의 수가 3개인 모형은 원형 발생과정에서 확률 파라미터 계산 시 오류로 만들어지지 않았으며 스트림의 수가 1개인 모형은 인식률이 상당히 낮았다. 스트림의 수가 2개인 모형이 전체적으로 인식률이 높았다. 개별 음소 모형에서 인식률은 학습 데이터에 대하여 평균 94.26%의 인식률을 그리고 시험 데이터에 대해서는 평균 82.44%의 인식률을 나타내었다. 그 중 각 스트림당 혼합의 수가 4개인 모형이 가장 좋은 모형인 것으로 나타났다. 집단 음소 모형 실험 또한 개별 음소 실험에서 높은 인식률을 기록한 HMM 모형에서 높은 인식률을 기록하였다. 그리고 학습 데이터에 대한 인식률은 상태수가 5인 모형에서 평균 93,88%을 상태수가 6인 모형에서는 평균 91.18%로 나타났으며 시험 데이터에 대한 인식률은 상태수가 5인 모형에서는 평균 82.15% , 상태수가 6인 모형은 평균 79.47%로 나타나서 상태수가 5인 모형이 더 나음을 알 수 있다. 단어 인식 실험에서는 학습데이터에 대해 평균 98.45%의 인식률을 그리고 학습데이터에 대해서는 평균 91.32%의 인식률을 기록하였다. 단어 인식에서의 인식률은 기존의 단어 인식기에서 나

타난 인식률 보다 약간은 낮지만, 음소 모형의 수를 좀 더 늘린다면 인식률이 상당히 개선될 것이라 예상된다. 본 연구에서는 데이터 획득 작업을 제외한 모든 작업을 동시에 수행할 수 있고 향후 개선 가능하도록 설계하였다.

본 연구에서는 제한된 화자의 수와 음성 데이터 및 모형에 대하여 동일 원형을 적용하여 실험하였으며, 실험후의 각각의 모형의 평가 작업을 통하여 최종 모형을 산출하였다. 그러나 본문에서 만들어낸 모형의 적용은 연속 숫자음에 국한되었고, 그 이외의 음성데이터 인식에 대한 적용은 시도하지 않았다. 한국어 음소에 대한 모형을 정확히 만들어내기 위해서는 실질적으로는 보다 많고 다양한 음성 데이터와 화자 및 서로 다른 원형을 적용하여 모형을 얻어내는 것이 바람직하다. 이 작업은 많은 인력과 비용이 필요하나, 자동분절 기법을 적절히 이용한다면 작업량을 크게 줄일 수 있다. 그 외에 동일 체계에서 원형이 서로 다른 HMM을 중복하여 사용하는 방법은 또 다른 개선책이 될 것이다.

Ⅵ 참고 문헌

[1] L. R. Rabiner and B. H. Juang, "An Introduction to Hidden Markov Models," IEEE ASSP Mag., pp. 26~40, July, 1990

[2] L. R. Rabiner, "A Tutorial on Hidden Markov Model and Applications in Speech Recognition, " Proc. IEEE. vol. 77, No.2, pp. 257~285, Feb. 1989

[3] G. D. Forney, "The Viterbi algorithm," Proc. IEEE, vol. 61, pp. 268~278, Mar. 1973.

[4] H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimazation for Spoken Word Recognition," IEEE Trains. on ASSP, vol .26, pp. 43~49, Feb, 1978

[5] N, Morgan and H. A. Bourlard, "Neural Networks for Statistical Recognition of continuous Speech," Proc. IEEE, vol. 83, no. 5, pp 742~770, May 1995

[6] T. Takara, K. Higa and I. Nagayama "Isolated Word Recognition Using HMM Structure Selected by the Genetic Algorithm," Proc. ICASSP 97 Munich, pp. 967~970

[7] L. Fissore, F. Ravera and P. Laface, "Using Word Temporal Structure in HMM Speech Recogonitoin," Ibid, pp. 975~978

[8] T. Svenson and F. K. Soong. "On the Automatic Segmentation of Speech Signals," Proc. ICASSP 87 vol. 1, pp. 77~80. Dallas. 1987

[9] R. A. Obrecht. "A New Statistical Approach for Automatic Segmentation of Continuous Speech Signals, IEEE. trans. on ASSP, vol. 36, no. 1, Jan. 1988

[10] R. A. Cole and Lily Hou, "Segmentation and Borad Classification of Continuous Speech," ICASSP-88, s10.12 New York, 1988

[11] J. W. Picon, "Signal Modleing Techniques in Speech Reconition," Proc. IEEE, vol. 81, no. 9, pp. 1215-1247, Sept. 1993

[12] S. E. Levinson, A. Ljolje, and L. G. Miller, "Large vocabulary speech recognition using a Hidden Markov model for acousitic/phonetic classification," ICASSP 88, 1988, pp. 505~508

[13] 이의천, 이강섭, 김순협, "음성 신호의 음소단위 구분화에 관한 연구," 음향학회지, 제 10권 4호, pp. 5-10, 1991

[14] 임완수, 임슬기, 이태호, "은닉 마르코프 모형을 이용한 한국어 자음의 인식에 관한 연구," 울산대학교 공학연구 논문집 25권 1호, pp. 55~65, 1994년 4월

[15] 김석수, "HMM을 이용한 자동 음성분절에 관한 연구," 울산대학교 공학석사 학위 논문 1996년 6월

[16] S.J.Young et. al "Hidden Markov Toolkit V1.5", Cambirdge University, Dec. 1997