정보기술응용연구
제1권 제3·4호
1999년 12월

HMM에 의한 한국어음성의 자동분할 시스템의 구현에 관한 연구[1)]

김 윤 중*, 김 미 경*, 이 인 동**

요 약

본 연구에서는 HMM(Hidden Markov Model) 및 Levelbuilding 알고리즘을 이용하여 인식대상 음소열의 표본 집합(훈련패턴 집합)을 입력으로 하는 음성의 자동 분할 시스템을 구현하였다.

본 시스템은 자연스럽게 발음되어진 연결음 음성으로부터 표준 음소모델을 생성한다. 본 시스템의 구성은 초기화 과정, HMM학습과정 그리고 Levelbuilding을 이용한 분리 및 Clustering 과정으로 구성되어 있다. 초기화 과정에서는 제어 정보를 이용하여 훈련패턴 집합으로부터 초기 음소 집합 군을 생성한다. HMM 학습단계에서는 각 음소 집합으로부터 음소모델을 생성한다. Levelbuilding을 이용한 분리 및 Clustering 단계에서는 음소 모델과 제어 정보를 이용하여 훈련패턴들을 음소 단위로 분리하고, 분리된 후보 음소들을 Clustering하여 음소 집합 군을 생성한다. 음소모델의 구성에 변화가 없을 때까지 이 작업을 반복 수행하여 최적의 음소모델을 생성한다.

본 연구에서는 3개 이하의 숫자단어로 구성된 연결단어 음성 패턴을 대상으로 실험하였다. 연결단어에 대한 음소의 표준모델 생성과정에서 가장 중요한 처리인 훈련패턴의 자동분할 과정을 분석하기 위하여 각 반복과정에서 분리된 정보를 그래프로 도시화하여 확인하였다.

1) 본 논문은 '97 교내 학술연구비 지원에 의하여 수행되었습니다.
*) 대전산업대학교 전자계산학과
**) 한국인식기술(주)

1. 서 론

1.1 연구의 배경 및 목적

음성은 인간이 가장 보편적으로 편리하게 이용하고 있는 정보의 전달 수단이고 음성인식 기술과 음성합성 기술은 이 음성을 통하여 인간이 컴퓨터와 대화할 수 있게 해주는 도구로서 정보화의 진전과 더불어 그 필요성이 더욱 증대되고 있다.

1970년대 무렵부터 미국 일본등 선진국에서는 국가주도의 대규모 프로젝트 형태로 연구가 추진되어 왔다. 그 결과 수십 단어 정도를 화자독립으로 인식하거나, 화자종속으로 수만 단어를 인식할 수 있는 시스템이 상용으로 등장하기에 이르렀다. 그러나 자연스러운 발음의 연속음성을 인식하는 데는 아직도 많은 연구가 요구되고 있으며 대용량 어휘의 인식 시스템의 연구에 초점이
모아지고 있다.

대용량어휘를 대상으로 하기 위하여서는 음소단위 인식이 수행되는 음성음향학적 모델에 기반을 둔 음성인식방법이 채택되어야 한다. 음소단위가 인식대상이 되기 위해서는 정확한 음소정보를 얻을 수 있어야 하고 이를 위해서는 정확한 음성의 분할이 요구된다. 음성분할은 음성의 음향학적정보를 참고하여 전문가가 수동으로 수행할 수도 있다. 그러나 스펙트로그램 판독 및 청취평가 등을 반복하여야 하고 일관성을 유지하기가 매우 어렵다.

이러한 문제를 해결하기 위하여 음성분할을 자동으로 수행하는 연구가 진행된 바 있다[1-4]. 음성신호에서 추출된 음향학적 정보만으로 음성을 분할하는 음향학적 분할방법(acoustic segmentation)[1-3]과 입력음성과 이 음성의 신호가 가지고 있는 음소열의 정보를 이용하여 음성을 분리하는 방법[4-6, 8-10]으로 분류된다. 후자에 속하는 "분할 K-Means training procedure" 방법[8-10]은 자연스럽게 발음되어진 단어음성들로부터 표준모델을 생성한다. 이 방법은 영어 음성을 인식대상으로 하였으며 단어 단위를 고려하였다.

본 연구에서는 1) HMM 및 LevelBuilding을 이용하는 "분할 K-Means training procedure" 방법에 기초하고 있으며, 2) 한국어 음성을 대상으로 음소단위의 자동 음성분할 시스템을 구현하였다. 3) 한국어 음성 분할을 위한 파라메터들을 결정하였고, 4) 실험을 통하여 음소 분할 성능을 확인하였다.

1.2 연구 내용

본 연구에서는 자연스럽게 발음되어진 연결음 음성(훈련패턴)으로부터 표준 음소 모델을 생성한다.

본 시스템은 훈련패턴을 그에 포함된 음소 수만큼 등 간격 분리하여 초기 음소 집합을 구성한다. 훈련패턴 집합의 정보가 정의된 제어정보를 이용하여 Levelbuilding의 처리 속도를 현저히 개선할 수 있었다. 그리고 Levelbuilding 알고리즘이 가지고 있는 인식 기능과 음소 분리 기능을 이용하여 훈련패턴들을 분리하고, 분리된 음소들을 제어정보를 이용하여 Clustering한다. 본 시스템은 인식 기능을 수행하면서 인식 대상 패턴들로부터 표준모델을 생성하므로 인식 시스템에 바람직한 표준모델을 생성할 수 있다. 본 연구에서는 음성의 자동분할 과정을 분석하기 위하여 각 반복과정에서 분리된 정보를 그래프로 도시화하여 확인함으로써 본 알고리즘의 효율성을 증명하였다.

본 논문의 구성은 2장의 시스템의 개요에서 시스템의 전체적 구성을 설명하고, 3장의 초기화 과정에서는 초기화에 수행하는 초기 음소 집합 군 생성과정을 설명하겠다. 4장의 HMM 학습단계에서는 HMM의 평가 및 재 측정 과정을 설명하고, 5장에서는 LevelBuilding알고리즘의 정의와 훈련패턴을 분리하는 과정을 자세히 살펴보고, 6장의 실험 및 결과에서는 숫자 음성들의 분리결과를 도시화하여 설명하고자 한다.

2. 시스템 개요

2.1 용어의 정의

본 알고리즘의 기술을 위하여 다음과 같이 용어를 정의한다.

정의 1. N
N은 인식단위 개체 수로써 인식하고자 하는 음소 수이다.

정의 2. 훈련패턴 집합
훈련패턴 집합 Ps는

$$Ps = \{ps_0, ps_1, ps_2, \ldots\} \text{ (식 2.1)}$$

와 같이 ps_m의 집합이다.

정의 3. 훈련패턴

훈련패턴 ps_m은 훈련을 위하여 채집된 m번째 음성 데이터이며 표본 데이터의 Sampling값의 집합이다.

$$ps_m = \{ps_m^0, ps_m^1, ps_m^2, \ \ldots\ , ps_m^{T-1}\}$$

(식 2.2)

이 훈련 데이터 ps_m^i는 LPC(Linear Predictive Coding) 변환과 벡터 양자화 과정을 수행하여 정수 o_m의 집합

$$o_m = (o_m^0, o_m^1, \ \ldots\ , o_m^{T-1})$$

(식 2.3)

으로 변환된 상태에서 처리된다.

$$ps_m \equiv o_m$$

(식 2.4)

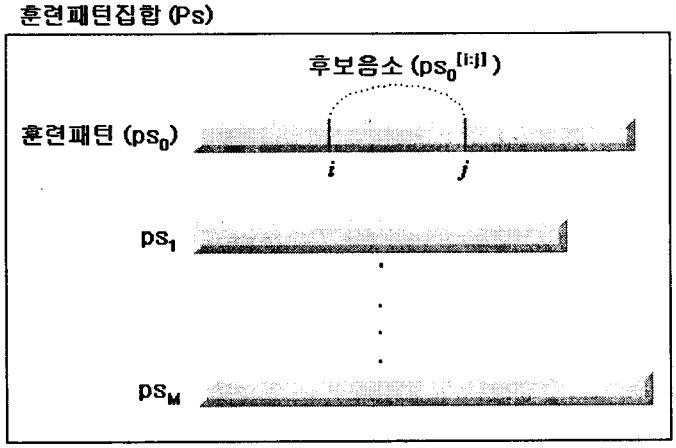
정의 4. 후보 음소

후보 음소는 훈련패턴의 부분집합으로, 훈련패턴에서 분리된 음소이다.

$$ps_m^{[i:j]} = \{ps_m^i, ps_m^{i+1}, ps_m^{i+2}, \ \ldots\ , ps_m^j\}, \quad j \geq i$$

(식 2.5)

아래의 그림은 훈련패턴집합과 훈련패턴 그리고 후보음소간의 포함관계를 나타낸 것이다.



[그림-2-1] 훈련패턴집합, 훈련패턴 그리고 후보음소간의 관계

정의 5. 음소 집합

음소 집합은 같은 음소이름을 갖는 후보 음소들로 구성된다.

$$w_s = \{w_s^0,\ w_s^1,\ w_s^2,\ \ldots\}$$

(식 2.6)

$$= \{\ ps_{m_0}^{[i_0:j_0]},\ ps_{m_1}^{[i_1:j_1]},,\ ps_{m_2}^{[i_2:j_2]}\ \ldots\ \}$$

$$w_s^1 = ps_{m_0}^{[i_0:j_0]}$$

정의 6. 음소 집합 군

음소 집합 군 W는 복수개의 음소 집합으로 구성된다.

$$W = \{w_0, w_1,\ ..\ , w_{N-1}\}$$

(식 2.7)

정의 7. 음소 모델

음소 모델은 음소 집합으로부터 생성되는 HMM 모델이다. i번째 음소 집합 w_i 로부터 생성되는 HMM 모델은 H_i 로 표기한다.

정의 8. 제어 정보

제어 정보란 훈련패턴 집합에 대한 정보로서, 각 훈련패턴의 패턴 이름, 음소 수, 음소 이름순으로 정의된다. 예를 들어 "영 2 ㅕ ㅇ"의 요소는 다음과 같은 의미를 갖는다.

- 영 : 훈련패턴 이름 (화일이름)
- 2 : 훈련패턴에 포함된 음소 수
- ㅕ : 음소 이름 1
- ㅇ : 음소 이름 2

중복 나열된 동일 음소는 하나의 음소로 취급한다. 예를 들어 "영이일"의 발음은 "ㅕㅇㅣㅣㄹ"과 같지만 제어 정보에서는 "ㅕㅇㅣㄹ"로 수록한다.

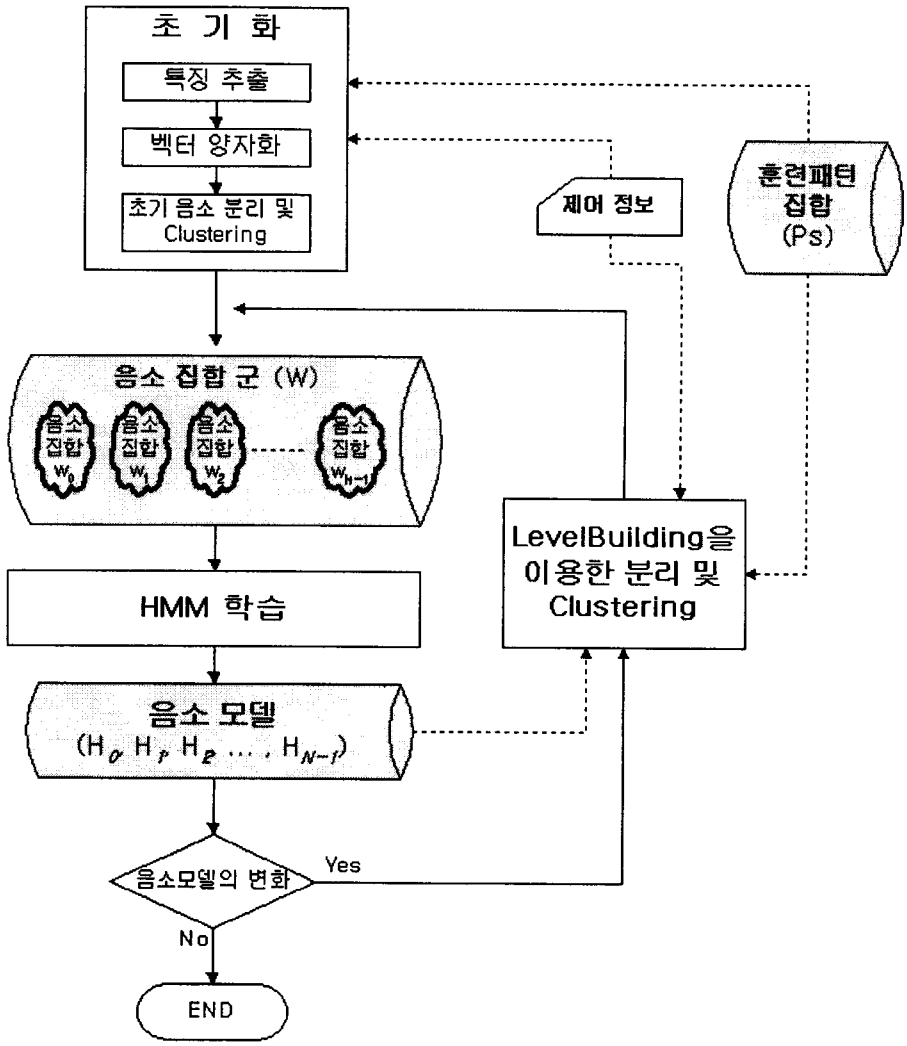
2.2 시스템의 개요 및 구성

본 시스템은 훈련패턴 집합 Ps를 입력으로 HMM 및 Levelbuilding 알고리즘 그리고 제어 정보를 이용하여 최적의 표준 음소모델을 생성하기 위한 음성 자동

분할 시스템이다.

본 시스템은 눈으로 분리해 내기 힘든 연결단어 음성패턴 상의 음소들을 Levelbuilding 알고리즘을 이용하여 분리해 내고, 분리된 음소들을 각 음소 집합으로 Clustering된다. 그리고 각 음소 집합으로 HMM을 학습시켜 음소모델을 생성한다.

본 시스템의 구성은 [그림-2-2]와 같다. 초기화 과정에서는 wave화일 형태로 구성된 훈련패턴 ps_i들의 특징을 추출하고, 벡터 양자화 한 후 제어 정보의 음소 수만큼 등 간격으로 분리하여 초기 음소 집합 군 W를 생성한다. 다음 HMM 학습단계에서는 각 음소 집합 w_i의 후보 음소들로부터 HMM 음소 모델 H_i을 생성한다. Levelbuilding을 이용한 분리 및 Clustering 단계에서는 음소 모델 H_i 와 제어 정보 및 훈련패턴 집합 Ps를 입력으로 Levelbuilding 알고리즘을 수행하여 각 훈련패턴을 음소 단위로 분리한다. 그리고 분리된 음소들을 Clustering하여 음소 집합 군 W를 생성한다. 이러한 과정은 음소 모델에 변화가 없을 때까지 반복 수행하여 최적의 표준모델을 생성한다.


초 기 화
특징 추출
벡터 양자화
초기 음소 분리 및 Clustering
제어 정보
훈련패턴 집합 (Ps)
음소 집합 군 (W)
음소 집합 W_0
음소 집합 W_1
음소 집합 W_2
음소 집합 W_{N-1}
HMM 학습
LevelBuilding을 이용한 분리 및 Clustering
음소 모델
(H_0, H_1, H_2 ... , H_{N-1})
음소모델의 변화
Yes
No
END

[그림-2-2] 시스템 구성도

3. 초기화 과정

초기화 과정에서는 주어진 훈련패턴 집합 Ps 를 제어정보의 음소 수만큼 등간격으로 분리하여 초기 음소 집합 군 W를 생성한다.

초기화 과정은 특징추출 과정, 벡터양자화 과정, 초기 음소분리 및 Clustering 과정의 3가지 단계로 구성된다.

3.1 특징 추출 과정

음성 데이터의 특징을 추출하는 방법으로 가장 보편적인 방법은 Filter Bank와 LPC 이다[11-13]. 본 연구에서는 훈련패턴의 특징을 추출하기 위하여 LPC 방법을 사용하였다.

본 연구에서 Recording Parameter값은 Sampling Rate 11 KHz, mono, 16bit[13]으로 하였다. LPC의 Parameter값으로 프레임의 길이는 300[11]으로 하였고, Shift Rate를 30으로 결정하여 실험하였다. Shift Rate를 큰 값으로 설정할 경우 발음되는 길이가 아주 짧은 초성 음성은 뒤에 이어져 발음되는 음소와의 식별이 불분명한 결과를 보였다. 예를 들면 숫자음성 중에서 "일"이라는 음성은 모음 "ㅣ"와 종성 자음"ㄹ"의 조합으로 이루어져 있다. 모음 "ㅣ"가 발음되는 시간은 아주 짧기 때문에 Shift Rate를 크게 설정할 경우 뒤 이어 발음되는 "ㄹ"에 묻혀 버리는 결과를 초래할 수 있다. Shift Rate를 100, 50, 30, 20으로 각각 설정하여 실험한 결과 Shift Rate를 50 이상으로 설정할 경우에는 짧게 발음된 음소의 식별이 불분명하였고 Shift Rate를 30 이하로 설정할 경우 식별이 분명하여 Shift Rate를 30으로 설정하였다.

3.2 벡터 양자화 과정

벡터 양자화 과정에서는 2차원 Vector로 구성된 훈련패턴들을 입력으로 하여 군집화(Clustering) 및 양자화를 거쳐 정수 열로 구성된 훈련패턴 집합을 생성한다.

군집화 과정에서는 MKM(Modify K-Means) 알고리즘을 이용하여 코드 북(CodeBook)을 생성한다. 양자화 과정에서는 군집화에 의해 생성된 코드 북(Codebook)을 이용하여 2차원 벡터의 훈련패턴 집합으로부터 정수 열로 구성된 훈련패턴 집합을 생성한다.

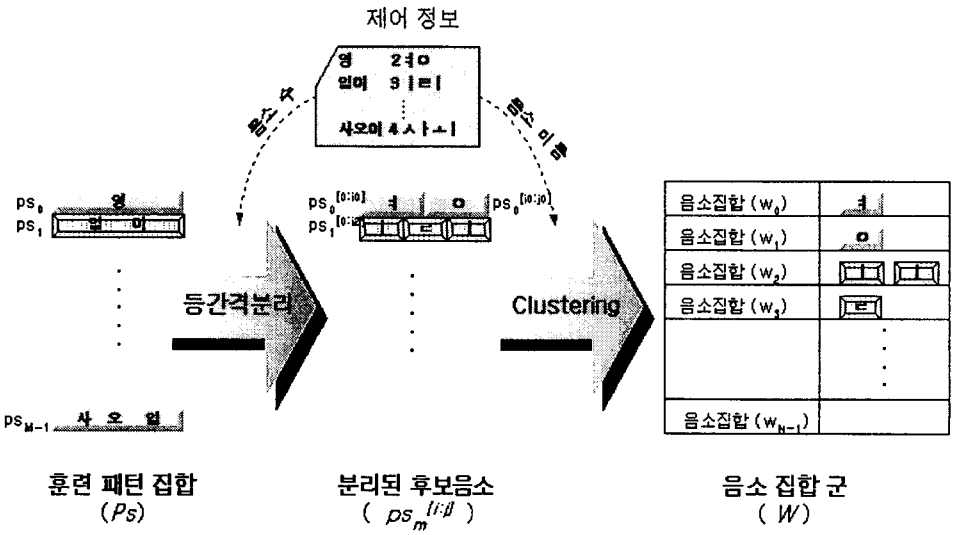
본 연구에서는 군집화의 parameter값인 Cluster 수를 15(실험데이터인 숫자음성에 포함된 음소 의 총 수)에서 25까지 설정하여 각각의 경우를 실험해 보았다. 실험 결과 Cluster 수를 21 이하로 설정할 경우 서로 다른 음소의 요소 값들이 같은 Cluster로 판단되는 오류가 나타났다. Cluster 수를 22 이상으로 설정할 경우 서로 다른 음소에 대한 식별이 분명하게 나타났다. 그래서 본 시스템에서는 군집화의 Cluster 수를 22로 결정하여 실험하였다.

3.3 초기 음소분리 및 Clustering 과정

초기 음소분리 및 Clustering 과정에서는 훈련패턴을 그에 포함된 음소 수로 분리하여 초기 음소 집합 군 W를 생성한다.

[그림-3-1]에서와 같이 각각의 훈련패턴을 그에 포함된 음소 수만큼 등 간격으로 분리한다. 그리고 분리된 음소들을 제어정보의 음소 이름에 대응되는 음소 집합으로 Clustering하여 초기 음소 집합 군을 생성한다.

[그림-3-1]에서는 초기 훈련패턴 분리과정을 보여주고 있다. "훈련 패턴 집합 "은 복수개의 훈련패턴(볼록한 직사각형)으로 구성되어 있다. 훈련패턴 상의 글자들은 각 훈련패턴에 포함된 음성들이다. 제어 정보의 음소 수는 훈련패턴을 등 간격으로 분리할 때 참조되고, 제어 정보의 음소 이름들은 분리된 후보 음소들을 Clustering 할 때 참조된다. "음소 집합 군"은 음소집합 w_i ($i = 0, 1, 2, 3, \ldots\ldots, N-1$)로 구성된다. "음소 집합 w_0"은 0번째 음소("ㅋ")의 음소 집합이고, "음소 집합 w_{N-1}"은 N번째 음소 집합을 나타낸다. 참고로 음소 집합 군의 Index는 0부터 시작한다.



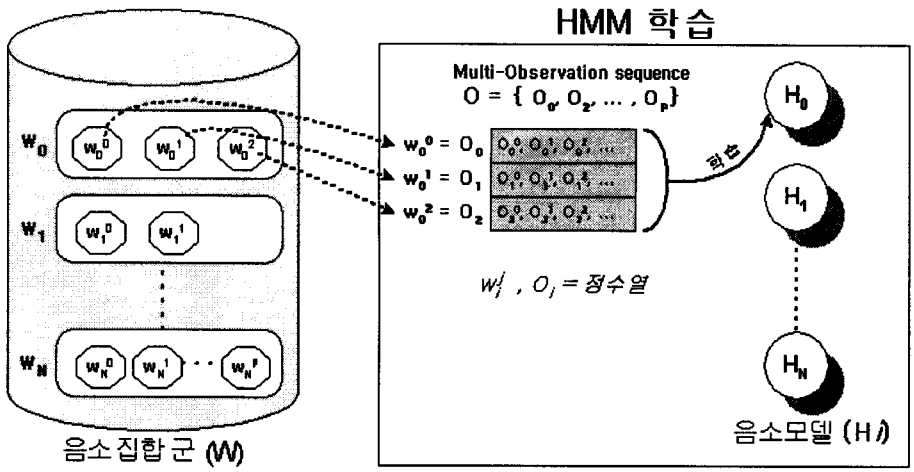
[그림-3-1] 초기 훈련패턴 분리과정

4. HMM 학습 과정

HMM 학습단계에서는 음소 집합 군 W 를 입력으로 하여 N개의 음소모델 H_i $(i=0,1, \ldots ,N-1)$을 생성한다.

본 연구에서는 DHMM의 복수 관측열(Multiple Observation sequence) 학습 알고리즘[11]을 사용하였다. HMM 음소 모델들은 각 음소 집합으로 학습(Reestimation)되어진다. HMM의 parameter인 상태 수는 5로 결정하고, 관측 심볼 수는 22로 결정하여 실험하였다.

음소 집합 군 W의 각 음소집합 $w_{i\,(i=0,1, \ldots ,N-1}$은 [그림-4-1]과 같이 Multiple Observation sequence O가 되며, 참고논문[11]의 방법으로 HMM을 학습을 시켜 음소 모델을 생성한다.



[그림-4-1] 음소집합과 HMM 학습

5. Levelbuilding을 이용한 분리 및 Clustering

Levelbuilding을 이용한 분리 및 Clustering과정에서는 음소 모델과 제어정보를 이용하여 훈련패턴 집합 Ps를 음소 단위로 분리하고, 분리된 음소들을 Clustering하여 음소 집합 군 W를 생성한다.

훈련패턴 ps_i는 음소 모델 H_i 및 제어정보를 이용하여 음소 단위로 분리된다. 입력 훈련패턴의 음소 분리점 b_i들이 처리가 된 후에 더 정확한 음소 분리점 b_i'가 구해진다. 분리점 b_i'로 분리된 후보 음소들은 제어 정보의 음소 이름에 대응되는 음소 집합 w_i로 Clustering되어 음소 집합 군 W를 생성한다.

[그림-5-1]은 3개의 음소로 구성된 훈련패턴이 본 알고리즘에 의하여 분리되는 과정을 보이고 있다.

훈련패턴 O를

$$O=(o_1, o_2, \ \ldots\ , o_{T-1})$$

T : Observation(훈련패턴의 요소) 수, 요소의 Index는 0부터 시작한다. 이라 하고, 이 훈련패턴이 세 개의 Segment

$$O=(S_0, S_1, S_2)$$

$$S_i=(o_{b_i}, \ \ldots\ , o_{b_{i+1}}) \equiv S(b_i, b_{i+1})$$

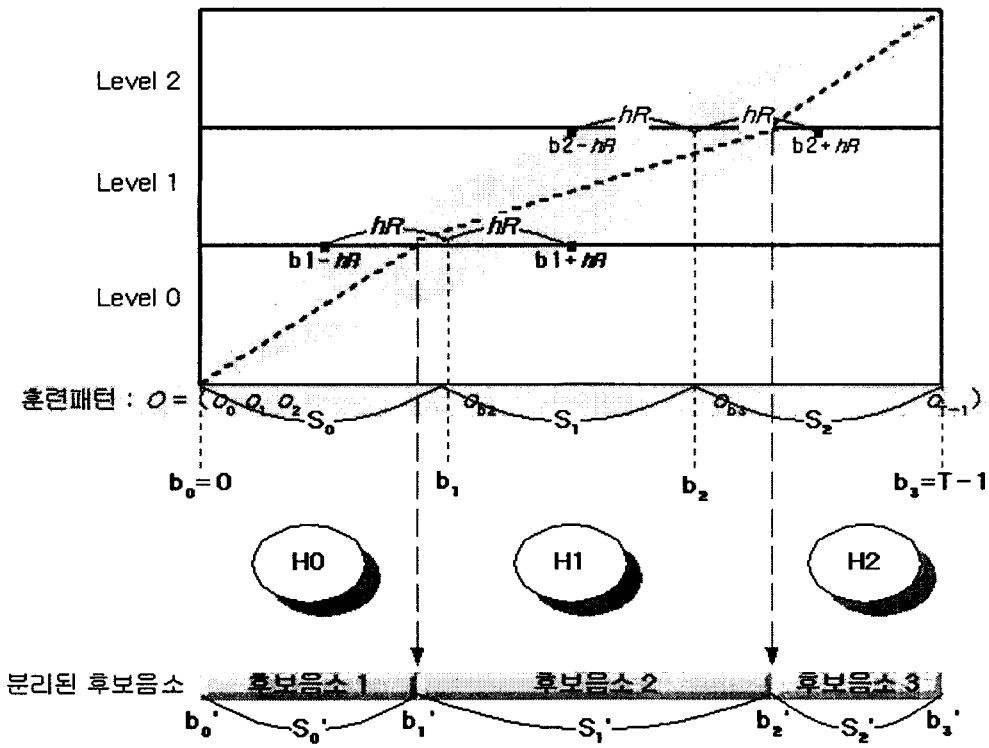
으로 구성되어 있다고 하자. 여기서 b_i는 i번째 Segment의 시작점(분리점) 인덱스이다. 각 Segment는 분리된 후보 음소 구간을 의미하며, 각 후보 음소의 HMM 모델은 H_0, H_1, H_2 로 한다.

본 알고리즘은 각 Segment의 새로운 분리점 b_i'를 구하는 것이 목적이다. 즉, 분리될 Segment $S(b_{i-1}, b_i)$에 대한 음소 모델 H_i의 확률들의 합이 최대가 되는 분리점 b_i'를 반복하여 계산하는 것이다.

b_i'를 계산하는 방법은 다음과 같다.

$$b_i' = \arg \max_{\overline{b_i}} \sum_{i=1}^{n} P(S(b_{i-1}, b_i) \mid H_i)$$

b_i'는 분리점 b_i 로부터 새로 계산되는 분리점이다. b_i'의 후보 영역 $\overline{b_i}$는 $b_i - hR \le \overline{b_i} \le b_i + hR$ 이다.



[그림-5-1] 음소 분리 과정

분리점의 계산 알고리즘은 다음과 같은 값들이 이용된다.

$\delta_n(j)$: n 번째 Segment에서 끝점이 j번째 Observation인 Observation 열에 대한 확률 값과 $\delta_{n-1}(i)$의 확률 값을 더한 누적 확률 값이다.

$P(i,j\,|H_n)$: i 번째에서 j번째까지의 Observation 열에 대한 확률 값으로 $P(i,j\,|H_n) \equiv P(\,(o_i, o_{i+1}, o_{i+2}, \ldots, o_j)\,|H_n) \equiv P(S(i,j)\,|H_n)$ 이다.

$\psi_n(j)$: $\delta_n(j)$를 최대화시키는 $i-1$번째 Segment의 분리점이다.

1번 초기화 과정에서는 후보 Segment $S(0,j)$에 대한 음소 모델 H_0의 확률 값을 $\delta_1(j)$에 저장한다. $\psi_1(j)$는 0을 저장한다.

2번 반복과정에서는 n번째 후보 Segment $S(i,j)$에 대한 음소 모델 H_n의 확률 값과 n-1 번째 확률 값 $\delta_{n-1}(i)$의 합이 최대가 되는 확률 값을 $\delta_n(j)$에 저장한다. 이때의 i 값은 $\psi_n(j)$에 저장된다.

3번 종료에서는 마지막 후보 Segment S_{NS}의 확률 값과 전(前) Segment S_{NS-1}까지의 확률 값 $\delta_{NS-1}(i)$의 합이 최대가 되는 값을 $\delta_{NS}(T-1)$에 저장한다. 이때의 i 값은 $\psi_{NS}(T-1)$에 저장된다.

4번 분리점 찾기에서는 마지막 분리점 $b_{NS}' = T-1$, 나머지 분리점들은 순환적으로 $b_n' = \psi_{n+1}(b_{n+1}')$으로 계산된다.

1) 초기화

$$\delta_1(j) = P(0,j\,|H_0) \qquad \text{(식 5.1a)}$$

$$\psi_1(j) = 0, \qquad b_1 - hR \le j \le b_1 + hR \qquad \text{(식 5.1b)}$$

2) 반복 과정

$$\delta_n(j) = \max_{b_{n-1}-hR \le i \le b_{n-1}+hR} [\delta_{n-1}(i) + P(i,j\,|H_n)] \qquad \text{(식 5.2a)}$$

$$\psi_n(j) = \arg \max_{b_{n-1}-hR \le i \le b_{n-1}+hR} [\delta_{n-1}(i) + P(i, j \mid H_n)], \quad 2 \le n \le NS-1,\ b_n - hR \le j \le b_n + hR \qquad \text{(식 5.2b)}$$

3) 종료

$$\delta_{NS}(T-1) = \max_{b_{NS-1}-hR \le i \le b_{NS-1}+hR} [\delta_{NS-1}(i) + P(i, T-1)] \qquad \text{(식 5.3a)}$$

$$\psi_{NS}(T-1) = \arg \max_{b_{NS-1}-hR \le i \le b_{NS-1}+hR} [\delta_{NS-1}(i) + P(i, T-1)] \qquad \text{(식 5.3b)}$$

4) 최적의 음소 분리 점 경로 찾기

$$b_{NS}' = N-1 \qquad \text{(식 5.4a)}$$

$$b_n' = \psi_{n+1}(b_{n+1}'), \qquad n = NS-1,\ NS-2,\ \ldots,1 \qquad \text{(식 5.4b)}$$

6. 실험 및 결과

6.1 연구 개발 환경 및 실험

본 연구의 개발환경은 Pentium 200 MHz의 CPU, 64 MByte의 M.Memory, Windows NT 운영체제 하에서 Visual C++ 6.0으로 구현하였다.

알고리즘 구현에서 LPC, MKM, Vector Quantization, DHMM 알고리즘 등은 SRLib[14] 라이브러리를 사용하였다.

본 연구에서는 표준 음소모델 자동생성 알고리즘을 실험하기 위해서 3개 이하의 숫자단어("영", "일", "이", ... ,"구")로 구성된 연속 숫자 음을 훈련패턴으로 사용하였다. 본 시스템에서는 훈련패턴의 음소 분리 과정을 분석하기 위하여 음소 분리 정보를 각 반복과정마다 그래프로 도시화하여 확인하였다.

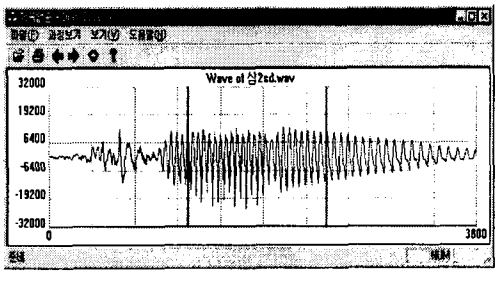
6.2 음소 분리의 예

본 논문에서는 실험 결과 중 한 예로 숫자 음 "삼"의 자동 음소 분리 과정을 그림으로 나타내었다.

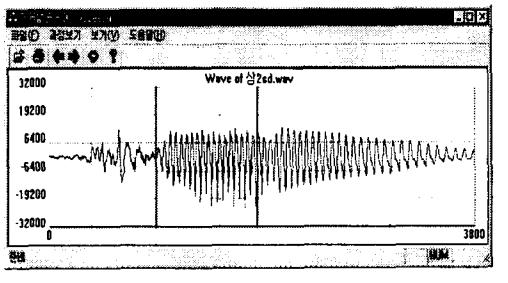
[그림-6-1] (a)에서 Y축은 Energy값의 수치를 표시하고 있고, X축은 시간 축을 나타낸다. 그림 중앙의 "삼1sd.wav"는 음성 패턴의 wave 파일 이름이다. 그리고 두꺼운 수직선은 훈련패턴에 포함된 음소와 음소 사이의 분리 선을 나타낸 것이다.

[그림-6-1] (a)는 초기화에 의해 등 간격으로 분리된 훈련패턴을 보여주고 있다. [그림-6-1] (a)에 표시된 "삼" 음성은 파찰음("ㅅ"), 유성음("ㅏ") 그리고 비음("ㅁ")으로 구성되어 있다. 음향학적 특성상 파찰음은 유성음과 비음에 비해 Energy값이 작다. 그러므로 파찰음 "ㅅ"의 Energy 값이 "유성음 "ㅏ" 보다 비교적 작다는 것을 짐작할 수 있다. [그림-6-1] (a)의 "삼" 음성에 포함된 음소를 발음되는 순서로 나열하면 "ㅅ ㅏ ㅁ" 순으로 구성된다. 여기서 우리는 "ㅅ"와 "ㅏ" 사이에 그리고 "ㅏ"와 "ㅁ"사이에 Energy값의 변화가 있음을 짐작할 수 있다. 실제로 예제 그림을 보면 오른쪽으로 가면서 Energy값이 커지는 것을 볼 수 있다. 이것은 파찰음 "ㅅ"에서 유성음인 "ㅏ"로 변해가면서 Energy값의 크기가 변화하는 것을 나타낸다.

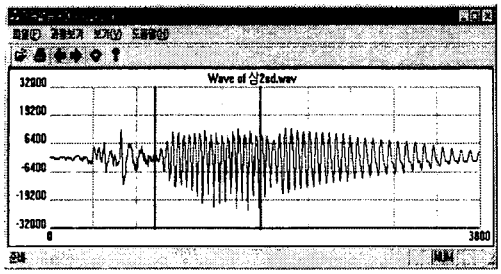
[그림-6-1] (a)와 [그림-6-1] (b)의 분리선을 보면 첫 번째 분리 선이 왼쪽 방향으로 이동해 가는 것을 볼 수 있다. [그림-6-1] (b)는 본 알고리즘을 1번 수행한 결과이고, [그림-6-1] (c)는 2번, [그림-6-1]의 (d)는 3번 수행한 결과이다. [그림-6-1] (e)는 본 알고리즘에 의해 분리된 최종 결과이다.



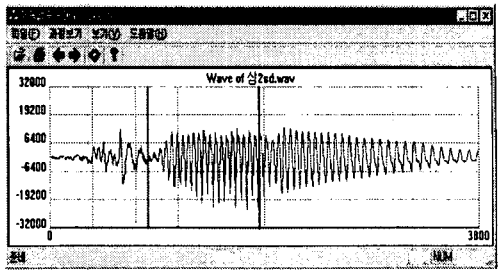
(a) 초기화에 의해 균등 분할된 훈련패턴



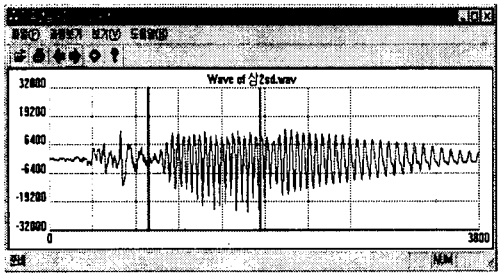
(b) Phase 1



(c) Phase 2



(d) Phase 3

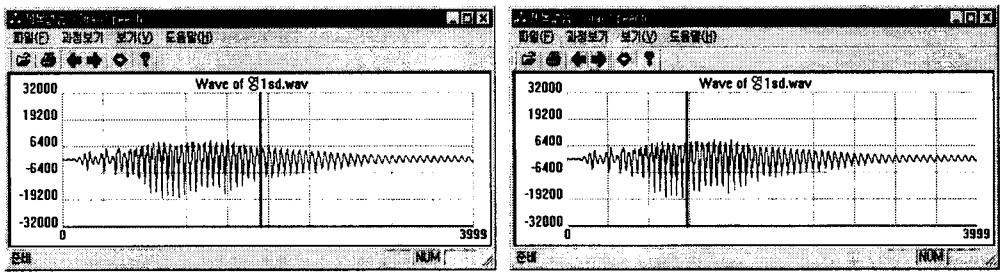


(e) 최종 분리된 결과

[그림-6-1] 음소 분리 과정의 예

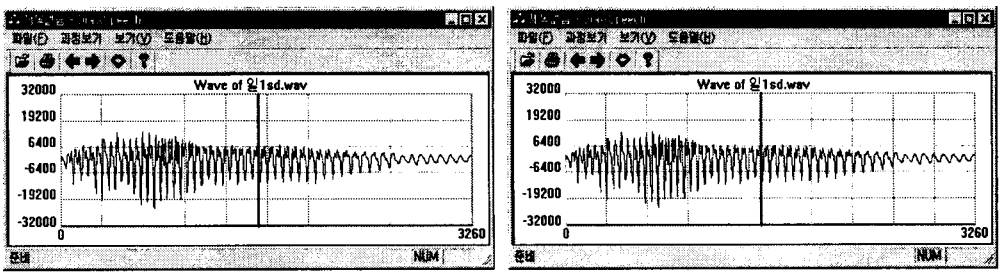
6.3 한국어 숫자 음성의 분리결과

[그림-6-2]에서 [그림-6-10]은 "영"에서 "구"까지의 숫자 음에 대한 음소분리 결과이다. 왼쪽 그림 (a)가 초기화에 의해 분리된 훈련패턴이고 오른쪽 그림 (b)가 음소 단위로 분리된 최종 결과이다. 단 하나의 음소로 구성된 숫자 음성은 분리되지 않기 때문에 실험 결과에서 제외하였다. 예를 들면 "이"는 모음 "ㅣ"만으로 이루어져 있고 "오"는 모음"ㅗ"만으로 구성되어 있다.



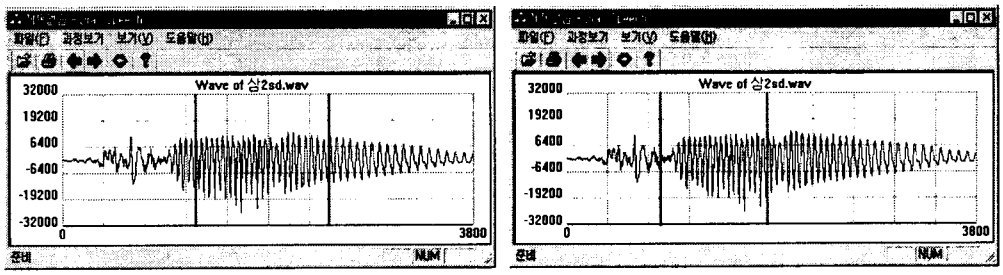
(a) 초기 음소 분리 (b) 음소 분리 결과

[그림-6-2] 숫자 음 "영"에 대한 음소 분리 결과



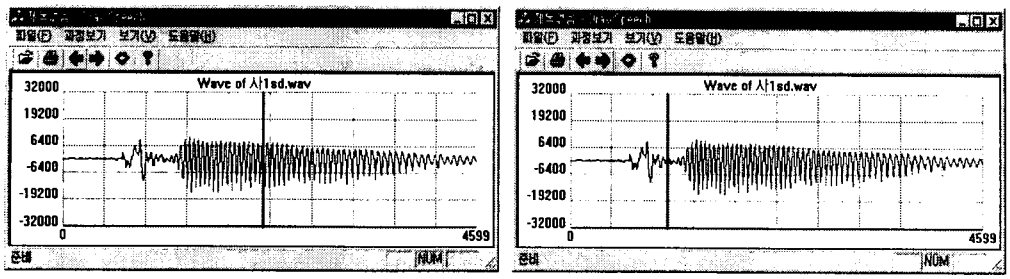
(a) 초기 음소 분리 (b) 음소 분리 결과

[그림-6-3] 숫자 음 "일"에 대한 음소 분리 결과



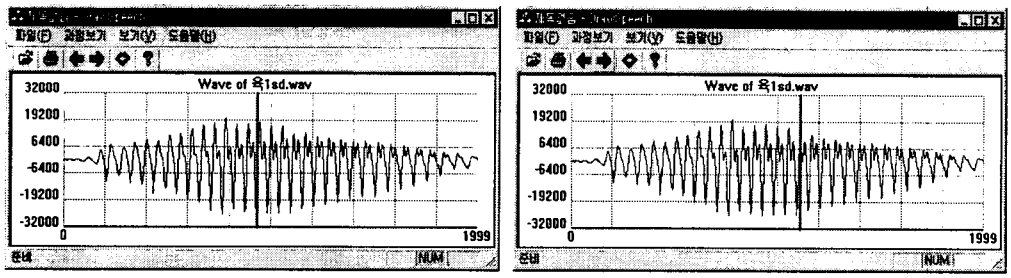
(a) 초기 음소 분리 (b) 음소 분리 결과

[그림-6-4] 숫자 음 "삼"에 대한 음소 분리 결과



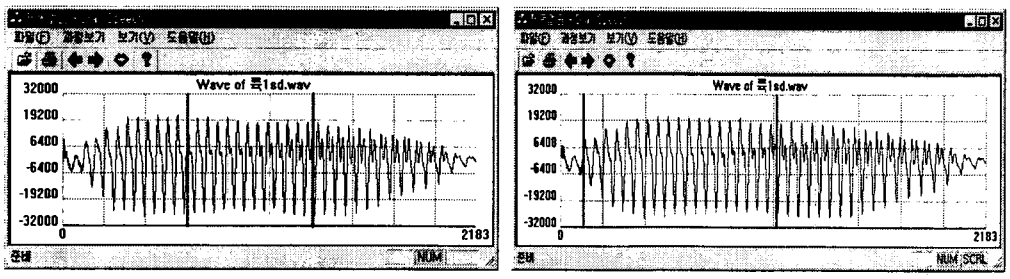
(a) 초기 음소 분리 (b) 음소 분리 결과

[그림-6-5] 숫자 음 "사"에 대한 음소 분리 결과



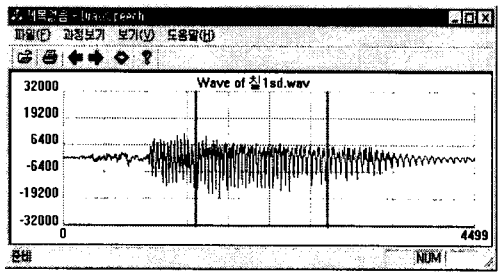
(a) 초기 음소 분리 (b) 음소 분리 결과

[그림-6-6] 숫자 음 "육"에 대한 음소 분리 결과

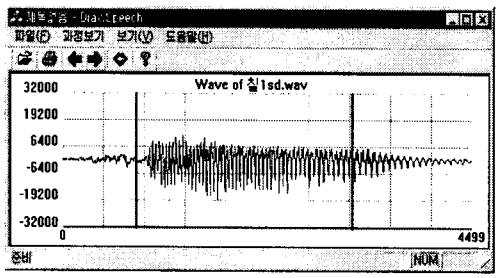


(a) 초기 음소 분리 (b) 음소 분리 결과

[그림-6-7] 숫자 음 "륙"에 대한 음소 분리 결과

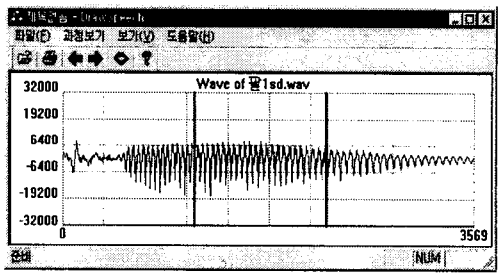


(a) 초기 음소 분리

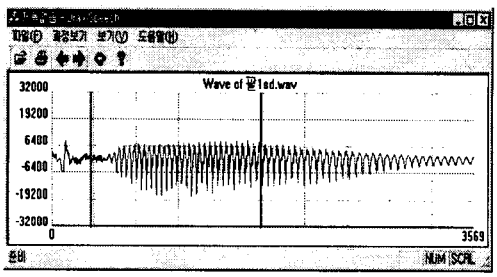


(b) 음소 분리 결과

[그림-6-8] 숫자 음 "칠"에 대한 음소 분리 결과

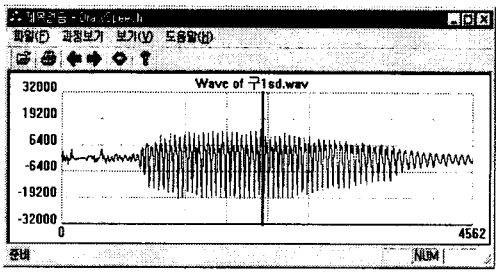


(a) 초기 음소 분리

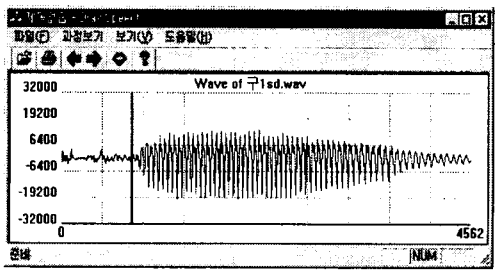


(b) 음소 분리 결과

[그림-6-9] 숫자 음 "팔"에 대한 음소 분리 결과



(a) 초기 음소 분리



(b) 음소 분리 결과

[그림-6-10] 숫자 음 "구"에 대한 음소 분리 결과

7. 결 론

패턴인식 방법에 기초한 음성인식 시스템에서 표본모델의 올바른 구축은 매우 중요하다. 본 연구는 "분할 K-Means procedure"에 기초하고 있으며, Levelbuilding 및 HMM 알고리즘을 이용하여 자연스럽게 발음된 한국어 음성으로부터 음성의 표본모델을 생성하기 위한 음성의 자동분할 시스템을 구현하였다.

본 알고리즘은 복수개의 훈련패턴으로부터 최적의 한국어 표준음소모델을 생성하기 위한 자동분할 시스템이다. 초기화 과정에서는 제어 정보를 이용하여 훈련패턴 집합으로부터 초기 음소 집합을 생성한다. HMM 학습단계에서는 각 음소 집합으로부터 HMM 음소모델을 생성한다. Levelbuilding을 이용한 분리 및 Clustering 단계에서는 음소 모델, 제어 정보 및 훈련패턴 집합을 입력으로 하여 훈련패턴들을 음소 단위로 분리하고 분리된 음소들을 Clustering하여 음소 집합군을 생성한다. 상기 과정을 반복하여 생성된 음소 집합 군으로 음소모델을 정련시킨다.

본 연구에서는 특징 추출과 벡터 양자화 그리고 HMM의 parameter값을 결정하였고, HMM과 Levelbuilding을 이용한 음성분할 알고리즘을 구현하였다. 또한 음성의 자동 분할 알고리즘을 한국어 음성을 대상으로 구현하여 숫자음에 대한 음소 모델을 생성하였다. 이러한 표준 음소 모델을 생성함으로써 Levelbuilding알고리즘을 이용한 인식 알고리즘을 쉽게 개발할 수 있다. 또한 개발 중인 모음 추출 시스템과 연계하여 모음-자음-모음, 모음-자음-자음-모음의 음소열 인식 시스템을 개발할 계획이다.

참고문헌

[1] T. Svendsen and F. K. Soong, "On the automatic segmentation of speech analysis," in Proceeding of International Conference on Acoustics, Speech and Signal Processing, pp. 77-80, Apr. 1987.

[2] J. R. Glass and V. W. Zue, "Multilevel acoustic Segmentation of Continuous Speech," in Proceeding of International Conference on Acoustics, Speech and Signal Processing, pp. 429-432, Apr. 1988.

[3] F Bimbot, G. Chollet, P. Deleclise and C. Montacie, "Temporal Decomposition and acoustic-phonetic decoding of speech," in Proceeding of International Conference on Acoustics, Speech and

Signal Processing, pp. 445-448, Apr. 1988.

[4] A. Ljolie and M. D. Riley, "Acoustic Segmentation and labeling of speech," in Proceeding of International Conference on Acoustics, Speech and Signal Processing, pp. 473-476, Apr. 1991.

[5] B. Whealtley, G. Doddington, C. Hemphill and J. Godfrey, "Robust Acoustic Time Alignment and Orthographic transcription with unconstrained speech," in Proceeding of International Conference on Acoustics, Speech and Signal Processing, pp. I-553-556, Apr. 1992.

[6] F. Brugnara, D. FalaVigna and M. Omologo, "Automatic Segmentation and labeling of speech based on hidden Markov models," Speech Communication, vol 12, no.4, pp. 357-370, Apr. 1993.

[7] 성종모, 김형순, 자동음성분할 시스템의 구현, 한국음향학회지, 제16권, 제5호, pp.50-59,1997.

[8] L. R. Rabiner, J. G.Wilpon, and B. H. Juang, "A Segmental K-Means Training Procedure for Connected Word Recognition," AT&T Tech. J., 65(3): 21-40, 1, May 1986.

[9] B .H. Juang and L. R. Rabiner, "The segmental K-Means Algorithm for Estimating Parameters of Hidden Markov Models", IEEE Trans. Acoustics, Speech, Signal Proc., vol.38(9) : 1639-1641, September 1990.

[10] L. R. Rabiner, J. G. Wilpon, and B. H. Juang, "A Model-Based Connected Digit Recognition System Using Either Hidden Markov Models or Templates,", Computer Speech, and Language, 1(2) : 167-197, December 1986.

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

[12] Atal, B. S. and Schroeder, M. R., "Predictive coding of speech signals," Proc. 6th Int. Cong. Acoust., C-5-4, 1968.

[13] 오영환, 음성언어정보처리, 홍릉과학출판사, p24-39, 1998.

[14] 김윤중, 김미경, 박은영, 음성인식 라이브러리(SRLib) 메뉴얼, 1998

A Study on the Implementation of an Automatic Segmentation System of Korean Speech based on the Hidden Markov Model

Yoon-Joong Kim, Mi-Kyoung Kim, In-Dong Lee

A Study on the Implementation of an Automatic Segmentation System of Korean Speech based on the Hidden Markov Model.

In this paper, We implemented an Automatic Segmentation System of Korean speech which is based on the Hidden Markov Model(HMM) and the level building algorithm. It segments each speech pattern into some phonetic units.

This system consists of three processes such as the initialization process, the HMM training process and the level building and clustering process. The initialization process generates a set of the initial phone group from the set of speech pattern. The HMM training process generates a set of the HMM of phoneme from the set of the phone group. The level building and clustering process segments the speech pattern into phonetic units using the phone model HMM and the level building algorithm, and then clusters the phonetic units into the phone group. The HMM training process and the level building and clustering process are processed repeatedly until positions of segmentations are not changed.

This system is experimented for the speech pattern which consists of less than 3 words. The result of segmented phoneme was depicted with the lines on the time domain wave of training data.

◆ 저자소개 ◆

김윤중 (Yoon-Joong Kim)

1981년 2월 : 충남대학교 전자공학과 (공학사)
1983년 2월 : 충남대학교 대학원 (컴퓨터 공학) 공학석사
1991년 2월 : 충남대학교 대학원 (컴퓨터 공학) 공학박사
1984년 8월 ~ 현재 : 대전산업대학교 교수
관심 분야 : 음성인식, 패턴인식

김미경 (Mi-Kyoung Kim)

1998년 2월 : 대전산업대학교 전자계산학과
현 재 : 대전산업대학교 전자계산학과 석사 과정

이인동 (In-Dong Kim)

1981년 2월 : 충남대학교 전자공학과 공학사
1987년 2월 : 충남대학교 교육대학원 (전자전공) 교육학석사
1991년 8월 : 충남대학교 대학원(전자 및 전산전공) 공학박사
현 재 : (주)한국인식기술 대표