Aurora DB를 이용한 잡음 음성 인식실험을 위한 Segmental K-means 훈련 방식의 기반인식기의 구현*

김희근(계명대), 정용주(계명대)

<차 례>

<Abstract>

An Implementation of the Baseline Recognizer Using the Segmental K-means Algorithm for the Noisy Speech Recognition Using the Aurora DB

Hee-Keun Kim, Young-Joo Chung

Recently, many studies have been done for speech recognition in noisy environments. Particularly, the Aurora DB has been built as the common database for comparing the various feature extraction schemes. However, in general, the recognition models as well as the features have to be modified for effective noisy speech recognition. As the structure of the HTK is very complex, it is not easy to modify the recognition engine. In this paper, we implemented a baseline recognizer based on the segmental K-means algorithm whose performance is comparable to the HTK in spite of the simplicity in its implementation.

* Keywords: Aurora DB, HTK, Segmental K-means algorithm, Baseline recognizer.

* 본 연구는 한국과학재단 목적기초연구(R01-2003-000-10242-0)지원으로 수행되었습니다.

1. 서 론

최근 음성인식 연구에서는 음성인식시스템의 강인성을 위해서 특징기법들 간의 성능비교에 대해서 관심이 높아지고 있다. 이를 위해서 많은 연구가들은 공통의 연구데이터로서 ETSI(European Telecommunications Standards Institute)에서 배포하는 Aurora DB를 사용하고 있다[1]. Aurora DB는 기본적으로 TI Digit 음성샘플에 부가잡음 및 채널왜곡을 가하여 만들어진 연속 숫자음 데이터이며, HTK 인식기를 바탕으로 기본적인 훈련절차를 제공하고 있다. 하지만, HTK를 이용하는 연구가들은 잡음음성인식을 위한 새로운 훈련방식에 대한 아이디어를 반영하기 위해서 HTK를 수정하는 것이 쉽지 않으며 시간비용도 많이 소모된다. 이러한 점에 기인하여 본 연구에서는 segmental K-means방식[3]에 기반한 훈련알고리듬을 Aurora DB에서 사용하는 방법을 고찰하였다. 이 방식은 이미 알려진 대로 HTK에서 모델 훈련을 위해서 사용되는 Baum-Welch 알고리듬[2]에 비해 구현이 간단하고, 수정이 용이하다. 두 방식의 비교 검증을 위한 구체적인 방법으로는, Aurora 2 DB에서 제공하는 훈련절차에 따라서 HTK를 이용한 경우와 제안한 segmental K-means 방식을 사용한 경우의 인식결과들을 상호 비교해 보았다. 본 논문의 구성은 다음과 같다. 먼저, 2장에서는 본 연구에 사용된 segmental K-means 방식에 대해서 간단히 소개하고 3장에서는 HTK를 이용한 훈련방법과 제안된 segmental K-means 훈련 방식간의 연관 관계에 대해서 설명한다. 그리고 4장에서는 HTK와 segmental K-means 방식과의 성능에 대해서 비교 검토하고 5장에서 결론을 맺는다.

2. Segmental K-means 방식의 개요

HTK에서는 HMM의 파라미터를 추정하기 위해서 Baum-Welch 알고리즘을 이용하지만, 본 연구에서는 파라미터의 추정방법으로 segmental K-means방식을 이용한다. Segmental K-means 방식은 Viterbi 알고리즘을 통해 발생하는 분할정보를 이용하여 HMM의 파라미터($\hat{\lambda} = \{\hat{c}_{jm}, \hat{\mu}_{jm}, \hat{U}_{jm}, \hat{a}_{ij}\}$)를 다음의 식과 같이 추정한다[3].

$$\hat{c}_{jm} = \frac{\text{혼합성분 m에 분류된 특징벡터들의 수}}{\text{상태 j에 할당된 전체 특징벡터들의 수}} \quad (1)$$

$$\hat{\mu}_{jm} = \text{상태 j , 혼합성분 m에 분류된 특징벡터들의 평균값} \quad (2)$$

$$\hat{U}_{jm} = \text{상태 j , 혼합성분 m에 분류된 특징벡터들의 공분산값} \quad (3)$$

$$\hat{a}_{ij} = \frac{\text{상태 i에서 j로의 천이횟수}}{\text{상태 i에서부터의 전체 천이횟수}} \quad (4)$$

여기서 $\hat{c}_{jm}$은 상태 j에 속한 각 혼합성분(mixture) m의 가중치(weight)이고 $\hat{\mu}_{jm}$은 상태 j, 혼합성분 m에 해당하는 평균벡터이며, $\hat{U}_{jm}$은 상태 j, 혼합성분 m에 해당하는 공분산 행렬(covariance matrix)이고 $\hat{a}_{ij}$는 상태 i에서 j로의 천이확률을 말한다. 잘 알려진 대로 위에서 보여주고 있는 segmental K-means방식의 계산식은 Baum-welch 알고리즘의 계산식보다 간단하여 쉽게 구현할 수 있다는 장점이 있다.

3. 제안된 훈련방식의 적용

3.1. HTK 훈련방법의 개요

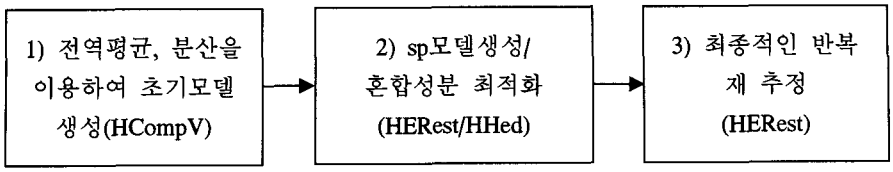
Aurora DB에서 제공하고 있는 HTK를 이용한 훈련 절차는 <그림 1>과 같으며, 이는 크게 3 가지 과정으로 다음과 같이 나누어 질 수 있다.

1) HTK의 툴(tool)인 HCompV로 전역 평균(global mean)과 전역 분산(global variance)을 구하여, 이를 전체 HMM 파리미터의 초기값으로 사용한다. 이러한 초기 모델 설정 방식을 flat-start라 한다[4].

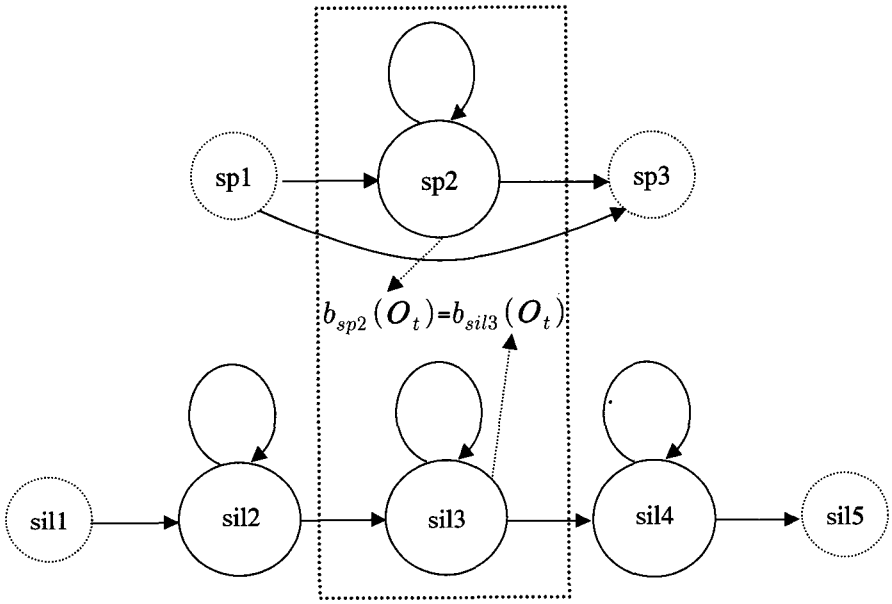
2) 초기모델 값이 정해진 다음에는 또 다른 툴인 HERest와 HHEd를 사용하여 반복 재 추정 및 HMM 개선(refine) 작업을 하여 목적한 HMM의 최종형태를 형성한다. 이러한 형성과정 중에 sp모델이 생성되고 각 HMM의 상태별 혼합성분의 개수가 최적화되어 설정된다.

3) 1),2) 단계를 거치면, HMM의 파라미터 값과 구조 등에 대한 기본적인 설정이 완성되며, 최종적으로 HERest를 이용하여 HMM의 파라미터 값을 반복 재 추정한다.

각 단어모델의 HMM 구조는 left-to-right 의 형태를 사용하며, 각 단어모델 별 상태의 개수는 묵음(silence)모델이 5개, sp모델이 3개, 나머지 숫자 모델들이 18개이다. 여기서 sp모델은 단어 간에 존재 가능한 묵음에 대한 모델이며, sp모델의 2번째 상태와 묵음 모델의 3번째 상태는 <그림 2>에서 보는 바와 같이 서로 공유된다(tied-state)[4].



<그림 1> Aurora 2 DB에서 제안된 훈련 흐름도

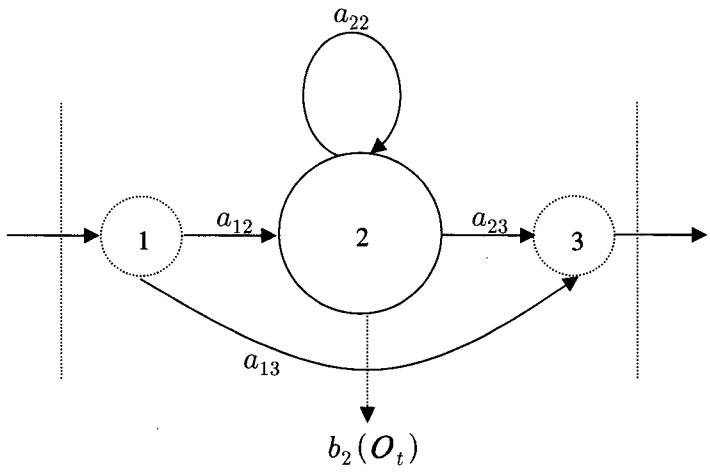


<그림 2> 서로 공유된 sp모델의 2번째 상태와 묵음 모델의 3번째 상태

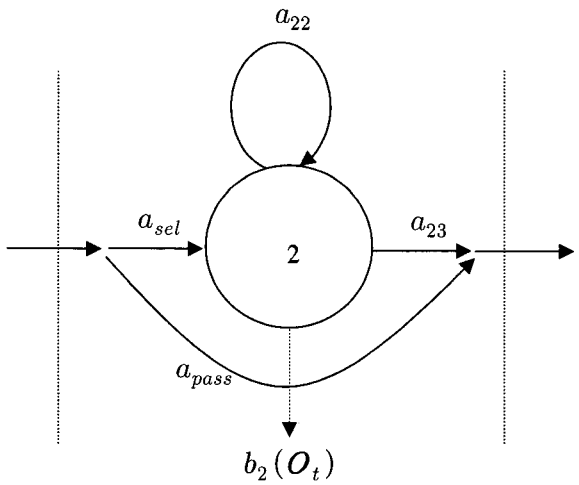
3.2. 제안된 segmental K-means방식

HTK의 Baum-Welch 훈련방법과 segmental K-means 훈련방식간의 성능을 비교하기 위해서는 3.1절에서 설명된 Aurora DB를 위한 HTK의 훈련절차를 segmental K-means 방식에 적용할 필요가 있다. 하지만, HTK의 훈련절차를 segmental K-means 방식에 그대로 적용할 때 약간의 문제가 발생한다. 그 첫 번째가 HTK에서 사용하는 HMM의 non-emitting 상태에 관한 것이다. Non-emitting 상태는 관측을 발생시키지 않는 상태로, 단어모델의 처음과 마지막 부분에 존재하여 여러 단어모델을 연결한 결합에서 특정 단어모델(sp)의 생략(skip)을 가능하게 해준다[4]. Segmental K-means 방식에서는 반드시 각 상태에 하나의 관측을 할당시켜야 하는데, non-emitting 상태는 관측을 발생시키지 않으므로 문제가 발생한다. 본 연구에서는 non-emitting 상태간의 천이가 허용되는 sp 모델에서 non-emitting 상태를 제거하는 대신에 간접적으로 sp 모델의 생략(skip)이 가능하도록 수정하였다. 따라서 segmental K-means 방식에서 사용한 전체 HMM 모델의 상태 개수는 HTK의 경우

에 비해서 두 개가 적게 된다. 한편, HTK의 sp 모델은 <그림 3>에 나타나 있으며 천이 a_{13}은 non-emitting 상태인 상태1과 상태3을 이용하여 sp 모델을 생략가능하게 해준다. 이러한 sp 모델은 segmental K-means 방식에서 다르게 나타나고 있는데 이는 <그림 4>에 나타나 있다.



<그림 3> HTK의 sp모델



<그림 4> 본 연구에서 제시된 sp 모델

<그림 4>의 sp 모델에서는 a_{sel}과 a_{pass}이라는 새로운 두 개의 가상의 천이확률을 사용한다. a_{sel}은 단어간의 묵음이 존재하여 sp 모델이 나타날 확률을 의미하

며, a_{pass}는 묵음이 존재하지 않아 sp 모델이 나타나지 않을 확률을 의미한다. <그림 3>의 천이와 <그림 4>의 새로운 천이 간에 관계를 역할에 따라 연관시킬 수 있는데, a_{12}는 sp 모델을 거치는 천이로 <그림 4>의 a_{sel}과 같고 a_{13}은 sp 모델을 건너뛰는 천이로 <그림 4>의 a_{pass}와 같은 역할을 한다. 훈련데이터에 대한 Viterbi 알고리듬기반의 분할정보를 이용하여, sp 모델의 새로운 천이 확률값 들은 아래식과 같이 추정된다.

$$a_{sel} = \frac{\text{sp 모델이 선택된 횟수}}{\text{훈련데이터의 전체 sp 모델의 개수}} \tag{5}$$

$$a_{pass} = \frac{\text{sp가 선택되지 않은 횟수}}{\text{훈련데이터의 전체 sp 모델의 개수}} \tag{6}$$

$$a_{sel} + a_{pass} = 1 \tag{7}$$

Aurora DB를 segmental K-means 방식에 적용시킬 때 나타나는 두 번째 문제는 공유상태(sp와 묵음의 경우)의 파라미터 값을 어떻게 추정하는가이다. 여러 가지 구현방법이 있지만 본 연구에서는 구현에 간단성에 주목하여 HMM의 파라미터들을 재추정할 때 서로의 통계 값들을 같이 더해주었으며, 그 방법은 수식(8), (9), (10)에 나타나있다.

$$\hat{c}_{jm} = \frac{n_{jspm} + n_{jsilm}}{N_{jsp} + N_{jsil}} \tag{8}$$

$$\hat{\mu}_{jm} = \text{묵음모델과 sp 모델에서 공유상태 j 의 혼합성분 m에 할당된 벡터들의 전체 평균} \tag{9}$$

$$\hat{U}_{jm} = \text{묵음모델과 sp 모델에서 공유상태 j 의 혼합성분 m에 할당된 벡터들의 전체 공분산값} \tag{10}$$

n_{jspm} = sp 모델의 공유상태 j 의 혼합성분 m에 분류된 벡터들의 수

n_{jsilm} = 묵음모델의 공유상태 j 의 혼합성분 m에 분류된 벡터들의 수

N_{jsp} = sp 모델의 공유상태 j 에 할당된 전체 벡터들의 수

N_{jsil} = 묵음모델의 공유상태 j 에 할당된 전체 벡터들의 수

세 번째 문제로는 일반적으로 많이 사용되는 flat-start를 통한 초기모델 설정 문제를 들 수 있다. Flat-start를 이용하여 초기모델을 구하게 되면 전역평균과 전역분산에 의해 전체 모델의 모든 상태가 동일한 파라미터 값을 가지게 된다. 이럴 경

우 Viterbi 알고리즘을 적용하면 유사도(likelihood) 값들이 전적으로 상태 천이확률에 의존하게 되므로, 분할 정보를 얻기 위한 Viterbi 알고리즘이 제대로 동작하지 않게 된다. 예를 들어, 모든 상태의 평균값과 공분산값이 동일하면 $b_1(O_t)$= $b_2(O_t)$=$b_3(O_t)$=... 가 되고 HTK에서 제안하는 초기모델의 상태 천이 확률값이 a_{jj}=0.6, a_{jj+1}=0.4이므로, Viterbi 알고리즘의 계산식에서는 항상 $b_1(O_t)a_{11}$ > $b_1(O_t)a_{12}$ 가 된다. 이로 인해 전체 관측들은 모두 처음 상태에 할당되게 되어 분할정보를 적절히 구할 수 없다. 그래서 segmental K-means방식의 훈련을 사용할 경우에는 초기값 설정을 위해서 일반적으로 균등분할(uniform segmentation)을 이용하지만, 인식실험 결과 이는 결과적으로 flat-start보다 성능이 떨어짐을 알 수 있었다.

4. 성능비교 실험

제안된 segmental K-means방식과 HTK 방식 간의 성능차이를 인식실험을 통해서 확인하였다. 인식실험에 사용된 Aurora 2 DB는 2개의 훈련모드(clean training, multi training)와 3개의 테스트 집합(set A, B, C)을 제공하고 있다[1]. <표 1>에는 제안된 segmental K-means 훈련방식과 기존의 HTK 방식에 의한 인식실험 결과가 나타나 있다. Segmental K-means 훈련방식을 이용한 경우에서는 초기 모델 설정을 위해서 균등분할을 이용하였으며, 초기모델 설정이후, <그림 1>에 나타나 있는 흐름도에 따라서 HMM 모델을 훈련하였다. <표 1>에서 알 수 있듯이 제안된 방식은 multi training 모드에서는 HTK와 비슷한 성능을 보이지만, clean training 모드에서는 HTK에 비해 성능이 다소 떨어진다. 이는 균등분할을 이용한 초기 모델 값이 HTK의 flat-start 방식에 비해서 훈련 데이터에 많이 의존하여 일반화가 부족하게 되어, 인식과 훈련데이터의 차이가 많은 경우에 인식성능의 많은 저하를 가져오는 것으로 생각된다.

<표 1> 균등분할을 이용한 초기모델 설정시의 segmental K-means 방식과 기존의 HTK 결과의 비교(word accuracy(%))

방법	Clean training			Multi training		
	Set A	Set B	Set C	Set A	Set B	Set C
Segmental K-means	59.18	55.28	62.72	87.46	86.51	83.16
HTK	61.34	55.75	66.14	87.82	86.27	83.77

초기모델의 보다 효율적인 설정을 위해서 우리는 segmental K-means 방식의 초기모델 값을 균등분할로부터 얻는 대신에 HTK 방식의 중간결과로부터 취하는 방안을 고려하였다. 여기서는 <그림 1>의 흐름도에서 2번째 과정까지 완성된 HTK 방식의 결과를 초기모델로 사용하였다. 이러한 초기모델의 재설정을 통한 segmental K-means 방식의 인식 결과가 <표 2>에 나타나있다. <표 2>에서 알 수 있듯이 초기모델 재설정의 효과는 매우 커서, 제안된 방식이 clean training 모드에서는 오히려 HTK 방식에 비해서 우수함을 알 수 있었다. 이러한 결과로부터 우리는 초기모델의 설정이 전체인식 성능에 미치는 영향이 매우 큼을 알 수 있었다. 또한, <표 2>의 결과로부터 우리는 제안된 segmental K-means 방식이 전반적으로 기존의 HTK 방식에 비해서 성능 저하가 거의 일어나지 않음을 확인할 수 있었다.

그리고 <표 3>과 <표 4>는 clean training 모드에서 두 가지 방식에 대한 set A의 결과를 보다 자세하게 제시하고 있으며, segmental K-means방식의 결과가 좀 더 좋음을 보여준다.

<표 2> 초기모델 값 재설정시의 segmental K-means방식의 결과
(word accuracy (%))

	Clean training			Multi training		
	Set A	Set B	Set C	Set A	Set B	Set C
Segmental K-means	62.54	57.02	66.86	87.54	86.30	83.08
HTK	61.34	55.75	66.14	87.82	86.27	83.77

<표 3> 초기모델 값 재설정시의 segmental K-means방식의 set A에 대한 보다 상세한 인식결과 (word accuracy(%))

Segmental K-means, Clean training, Multi-condition testing					
Set A					
	Subway	Babble	Car	Exhibition	Average
Clean	98.93	99.00	98.96	99.11	99.00
20	97.01	91.96	97.58	96.30	95.73
15	93.61	76.69	91.50	92.81	88.65
10	79.89	51.42	69.58	78.56	69.86
5	53.09	27.33	36.06	48.63	41.27
0	25.18	9.04	14.17	20.30	17.17
-5	10.87	1.15	9.31	9.35	7.67
Average (0dB-20dB)	69.77	51.29	61.78	67.32	62.54

<표 4> HTK방식의 인식결과 (word accuracy(%))

HTK, Clean training, Multi-condition testing					
Set A					
	Subway	Babble	Car	Exhibition	Average
Clean	98.93	99.00	98.96	99.20	99.02
20	97.05	90.15	97.41	96.39	95.25
15	93.49	73.76	90.04	92.04	87.33
10	78.72	49.43	67.01	75.66	67.71
5	52.16	26.81	34.09	44.83	39.47
0	26.01	9.28	14.46	18.05	16.95
-5	11.18	1.57	9.39	9.60	7.94
Average (0dB-20dB)	69.49	49.89	60.60	65.39	61.34

한편, 지금까지 우리는 주로 HTK와 제안된 방식간의 훈련방법 차이에 따른 인식성능을 비교하였는데, 이러한 차별의 원인이 두 방식의 훈련부분 외에 인식부분의 차이에 의한 것이 아닌가 하는 점을 확인하기 위해서 HTK 방식으로 최종 훈련된 HMM 모델을 제안된 방식에서 사용한 인식엔진을 통하여 인식실험을 해보았다. 이는 <표 5>에 나타나 있으며, 서로 간의 결과를 비교해보았을 때 큰 차이가 없음을 알 수 있었으며, 두 방식의 인식부분 사이에는 어느 정도 호환성이 있다고 생각된다.

<표 5> 제안된 방식(segmental K-means)과 HTK 방식 간의 인식부분 호환성 실험결과 (word accuracy(%))

	Clean training			Multi training		
	Set A	Set B	Set C	Set A	Set B	Set C
Segmental K-means	61.24	55.57	66.13	87.66	86.22	83.51
HTK	61.34	55.75	66.14	87.82	86.27	83.77

5. 결 론

본 연구에서는 Aurora DB를 위한 잡음음성인식에서 기존의 HTK에서 사용되는 Baum-Welch 훈련방식 대신에 segmental K-means 훈련방식을 적용할 것을 제안하였고, 구체적으로 다양한 인식실험을 통해서 제안된 방식이 그 구현의 간단성과 수정의 용이성에도 불구하고 Aurora DB에서 제시하는 기준결과와 거의 유사한 성능을 보임을 알 수 있었다. 따라서 제안된 방식을 이용함으로서 잡음음성인식을 위

한 새로운 훈련 알고리듬을 구현하기가 수월해졌을 뿐만 아니라, Aurora DB에서 제공하는 기준 결과와의 비교검토도 용이 할 것으로 생각된다.

참 고 문 헌

[1] D. Pearce, H. Hirsch, "The Aurora experimental framework for the performance evaluation of speech recognition systems under conditions", *Proc. ICSLP 2000,* vol. IV, pp. 29-32, Beijing, China.

[2] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", *Proc. IEEE,* vol. 77, pp. 257-286, Feb. 1989.

[3] B. H. Juang and L. R. Rabiner, "The segmental k-means algorithm for estimating the parameters of hidden Markov models", *IEEE Trans. Accoust., Speech, Signal Processing*, vol. ASSP-38, no. 9, pp. 1639-1641, Sep. 1990.

[4] S. Young, G. Evermann, T. Hain, et al., *The HTK Book* (for HTK version 3.2.1), http://htk.eng.cam.ac.uk/

[5] 김희근, 정용주, 배건성, "Aurora DB를 이용한 잡음 음성 인식실험에서 기반인식기의 구현에 관한 연구", 한국음성과학회, *제17차 한국음성과학회 추계학술대회*, pp. 77-87, 2005.

접수일자: 2006년 2월 15일
게재결정: 2006년 3월 20일

▶ 김희근(Hee-Keun Kim)
주소 : 대구광역시 달서구 신당동 1000, 계명대학교 (704-701)
소속 : 계명대학교 전자공학과
전화(Tel) : +82-53-580-5925
E-mail : hk2283@kmu.ac.kr

▶ 정용주(Young-Joo Chung) : 교신저자
주소 : 대구광역시 달서구 신당동 1000, 계명대학교 (704-701)
소속 : 계명대학교 전자공학과
전화(Tel) : +82-53-580-5925
E-mail : yjjung@kmu.ac.kr