한국음향학회지 제20권 제7호 pp. 55~62 (2001)

한국어 주소 음성인식의 고속화를 위한 적응 프루닝 문턱치 알고리즘

An Adaptive Pruning Threshold Algorithm for the Korean Address Speech Recognition

황 철 준*, 오 세 진**, 김 범 국*, 정 호 열**, 정 현 열**
(Chul-Jun Hwang*, Se-Jin Oh**, Bum-Koog Kim*, Ho-Youl Jung**, Hyun-Yeol Chung**)

*대구과학대학 정보전자통신계열, **영남대학교 전자정보공학부
(접수일자: 2001년 5월 21일; 채택일자: 2001년 9월 20일)

음성인식의 고속화를 위한 저자들에 의한 기존의 연구에서는 탐색이 진행함에 따라 시간방향의 탐색공간 문턱치를 가변적으로 적용하여 인식률의 저하없이 인식속도를 개선시켰다. 이 방법은 탐색 공간을 효과적으로 줄일 수는 있었으나 문턱치를 결정하기 위해서 여러 번의 사전 실험을 수행하여야 하는 번거러움이 있었다. 이러한 문제점을 해결하기 위하여 본 논문에서는 이전 탐색구간에 대한 최대우도와 후보들의 우도를 이용하여 현재 탐색구간의 문턱치를 탐색이 진행하는 과정에서 자동적으로 구하는 적응 프루닝 문턱치 알고리즘을 제안하였다. 제안한 알고리즘의 유효성을 확인하기 위해 국내 행정단위 시 (도), 구 (군), 동 (읍, 면), 번지를 구성하는 단어로 구성된 주소 인식 시스템에 적용하여 기존의 방법과 제안한 방법을 비교 검토하였다. 인식실험 결과, 연결단어 인식률 96.0%, 단어 인식률이 98.7%인 경우를 기준으로 하였을 때 제안된 방법이 기존의 고정 프루닝과 가변 프루닝 문턱치에 비하여 인식률 저하없이 각각 14.4%와 9.14%의 탐색 공간을 상대적으로 줄일 수 있어 제안된 방법의 유효성을 확인할 수 있었다.

핵심용어: 프레임단위 적응 프루닝, 가변 프루닝, 목구조형 사전, 주소입력시스템, One-Pass Viterbi 빔 탐색
투고분야: 음성처리 분야 (2.5)

In this paper, we propose a new adaptative pruning algorithm, which effectively reduces the search space during the recognition process. As maximum probabilities between neighbor frames are highly interrelated, an efficient pruning threshold value can be obtained from the maximum probabilities of previous frames. The main idea is to update threshold at the present frame by a combination of previous maximum probability and hypotheses probabilities. As present threshold is obtained in on-going recognition process, the algorithm does not need any pre-experiments to find threshold values even when recognition tasks are changed. In addition, the adaptively selected threshold allows an improvement of recognition speed under different environments. The proposed algorithm has been applied to a Korean Address recognition system. Experimental results show that the proposed algorithm reduces the search space of average 14.4% and 9.14% respectively while preserving the recognition accuracy, compared to the previous method of using fixed pruning threshold values and variable pruning threshold values.

Keywords: *Adaptive pruning, Variable pruning, Tree-structred lexicon, Address input system, One-pass viterbi beam search*

ASK subject classification: *Speech signal processing (2.5)*

책임저자: 정현열 (hychung@ynucc.yeungnam.ac.kr)
712-749 경북 경산시 대동 214-1
영남대학교 전자정보공학부
(전화: 053-810-2496; 팩스: 053-814-5713)

I. 서 론

컴퓨터 하드웨어 기술의 급속한 진보와 음성처리 기술의 발전으로 인하여 음성인식의 실용화가 실질적인 문제로서 관심이 증대되고 있다. 이러한 관심이 증대되면서, 음성인식에 관한 연구는 실용화에 초점이 모아지면서 최근 몇 년간의 눈부시게 발전하여 일부 태스크에서 상업용 시스템이 구현되고 있는 실정이다[1,2].

음성인식 시스템의 실용화를 위해서 가장 중요한 것은 높은 인식 성능을 가지면서 동시에 실시간으로 인식되어야 할 필요가 있으나, 이 두 요구조건은 상충되는 사항이다. 예를 들어, 인식시간을 줄이기 위해 탐색 공간을 대량으로 프루닝하면서 간단한 음향학적 모델을 사용하면 인식 속도를 쉽게 향상시킬 수는 있지만 이에 따르는 인식률의 저하는 피할 수 없다. 따라서, 현재까지의 연구결과에 의하면 고립 단어 단위에서는 인식률을 향상시키거나 혹은 그대로 유지하면서 인식 속도를 높이는 것에 대해서는 어느 정도 연구 성과가 있지만, 대용량 어휘를 대상으로 하는 연결음성인식 또는 연속음성인식에서는 아직까지 많은 연구가 필요하다. 실제로 이용할 수 있는 실용화 시스템을 구축하기 위해서는 높은 인식성능과 빠른 인식 속도의 두 조건을 동시에 만족하지 않으면 안 된다[3].

인식률의 경우 고립단어 인식에 있어서는 약간의 잡음이 있는 환경하에서도 95%이상의 인식 성능을 가지며, 한정된 태스크 범주내의 연속음성인식에서도 90%이상의 높은 인식률을 가진 시스템이 많이 개발되고 있으며, 인식 태스크를 확장하기 위한 여러 가지 연구들이 진행되고 있다[4-6].

인식시간의 경우, 국외에서는 Name Dialing System, 증권 거래 시스템 등과 같이 고립 단어를 대상으로 하는 수종의 시스템이 개발되어 실용화되고 있으며, Dragon Dictate, 날씨 안내 시스템 등과 같이 한정된 태스크에서의 연속음성 인식에서도 거의 실시간으로 동작하는 시스템이 많이 개발되어 실용화 단계에 있다[7,8]. 또한 자연발화 (Natural Speech) 인식에 대한 연구도 활발하게 진행 중에 있다. 국내의 경우에 있어서는 최근의 음성인식에 대한 관심의 증대로 인하여 증권 안내 시스템, 부서 안내 시스템 등과 같이 고립단어를 대상으로 하는 인식 시스템이 개발되어 실제 상용화되고 있지만, 대어휘를 대상으로 하는 실시간 음성인식 시스템 구현을 위한 고속화에 대한 연구는 아직까지 많이 부족한 실정이다[1,2].

따라서 저자들은 음성인식 시스템의 실용화를 위해 높은 인식 성능뿐만 아니라 실시간으로 동작하는 시스템을 구축하기 위하여 연구를 진행해 왔다[9-13]. 그 연구 결과로서 개발된 인식시스템은 범용 사운드카드가 장착된 개인용 컴퓨터의 윈도우 환경에서 동작하는 음성인식 기능을 가진 주소 인식시스템으로, 한국어 주소의 특징을 고려하여 연결단어 인식을 태스크로 하고 있다. 기존의 연구 결과에서 전국의 모든 주소명을 대상으로 연결단어 (연결 주소음성) 인식률 96.0%, 고립단어 인식률 98.7%을 달성하고 있다. 인식 시간의 경우, 컴퓨터 시스템의 성능에 따라 인식 시간 측정에 차이가 있을 수 있기 때문에 인식 시간을 추정할 수 있는 탐색공간을 계산하였으며, 가변 프루닝 문턱치 알고리즘을 적용한 결과 전체 탐색 공간 중 평균 28.87%의 공간탐색으로 인식이 수행됨을 확인하였다[12,13]. 이 연구 결과는 각 프레임에서 후보 단어들을 효과적으로 제한할 수 있음을 보이고 있지만, 여전히 불필요한 공간 탐색이 이루어지고 있음을 확인할 수 있었다.

본 논문에서는 이와 같은 불필요한 탐색 공간을 보다 효과적으로 제한하기 위하여 인식 과정 중에 탐색대상 공간을 자동으로 제한하여 실제 탐색공간을 감소시키는 방법을 모색하기로 한다. 이때 탐색공간 축소로 인한 인식성능 저하는 일어나지 않도록 하는 실시간 음성인식을 위한 고속화 알고리즘 개발을 연구의 대상으로 하고자 한다.

논문의 구성은 다음과 같다. Ⅱ장에서는 기존의 음성인식 고속화 알고리즘을 소개하고, Ⅲ장에서는 본 연구에서 제안하는 프레임 단위 적응 프루닝 문턱치 알고리즘에 대하여 설명하고, Ⅳ장에서는 전체 시스템의 구성과 인식실험 방법, Ⅴ장에서는 제안한 고속알고리즘을 이용한 인식실험 결과를 기술한 다음, 마지막으로 Ⅵ장에서 본 논문의 결론을 맺는다.

II. 음성인식 고속화 알고리즘

음성인식에 있어서 가장 간단한 방법은 예측되어진 전체의 후보와 입력음성을 비교하는 방법이다. 그러나 이 방법은 대상 어휘수가 증가하고 인식 알고리즘이 복잡해짐에 따라 대규모 탐색공간을 필요로 하고 이에 따라 많은 처리시간을 요구한다. 실시간 음성인식을 위해서는 전체의 후보와 정합을 행하지 않고도 고정도의 인식성능을 얻을 수 있는 효율적인 탐색수법이 필요하게 된다.

이하에 탐색 공간을 줄이기 위하여 전형적으로 사용하는 고속화 알고리즘과 기존의 연구에서 제안된 알고리즘을 간략한다.

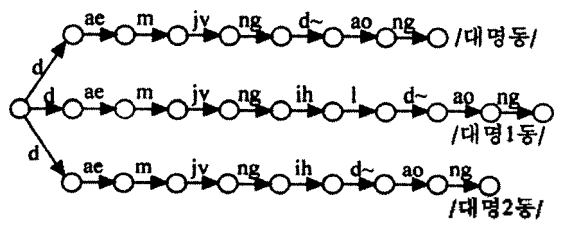
2.1. 목구조형 사전

일반적으로 탐색되어지는 상태 네트워크 (음소열)가 어휘내의 각 단어에 대해 선형적으로 결합되어 있기 때문에, 이러한 구성으로 인해 실제로 탐색되는 모델의 수는 어휘 수에 비례하게 된다.

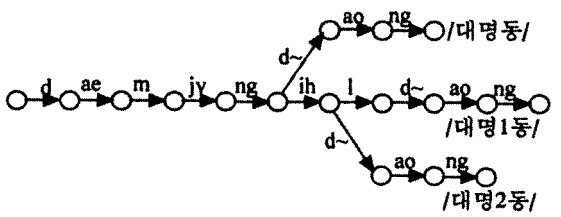
따라서, 탐색공간의 크기를 줄이기 위해서는 목구조형 사전 (Tree-structured lexicon)을 구성하는 방법을 도입할 필요가 있다[14,15]. 이것은 고속화를 위한 언어적 수법 중의 하나로써 많은 단어들이 어두부분에 동일한 접두어로 시작하고 있는 점에 착안하여 이를 공유하여 중복을 피함으로서 탐색공간을 줄이는 방법이다. 초기에 목구조는 여러 시스템에서 후보성분을 생성하기 위한 고속 정합에 사용되었다. 그림 1에서는 선형결합과 목구조형 사전 구성의 예를 나타낸다.

2.2. 고정 프루닝 문턱치

음성인식을 수행하기 위해서는 출력확률 계산과 탐색의 2가지 계산과정을 필요로 한다. HMM (Hidden Markov Model)을 이용한 음성인식에서의 출력확률 계산은 임의의 한 시점에서 관측된 음성을 출력하는 주어진 HMM (Hidden Markov Model)의 상태의 확률계산이며, 탐색은 주어진 음성 입력에 대한 최상의 상태열을 구하는 문제로 볼 수 있다. 이러한 탐색에 소요되는 시간은 음향학적 모델의 복잡성에 의해서는 크게 영향을 받지 않으나, 인식대상의 규모에 따른 영향은 크다. 즉, 인식에 있어서 모든 가능한 상태열들을 고려할 경우, 입력된 음성에 대한 최



(a) Linearly connected lexicon



(b) Tree structured lexicon

그림 1. PLUs를 이용한 한국어 단어사전의 예
Fig. 1. An example of lexicon for Korean word based on PLUs.

고 우도의 상태 열 (단어, 문장)을 찾기 위한 탐색공간은 지수 함수적으로 증가한다.

현재까지 대부분의 시스템에서는 프레임동기형의 빔 탐색법을 이용하고 있는데 이 방법은 각 후보의 우도를 비교하고 상위 일정 개수 (문턱치 이하의 것)에 대해서만 후속 정합을 고려하는 방법으로 다음과 같이 나타낸다[16].

$$D_{\min}(i,j) \le D_{\min}(i,j^*) + \delta \tag{1}$$

이 방법은 i 프레임에서의 최적의 경로 (i,j^*)에 대해 문턱치 δ 이내의 상위 몇 개 (빔 폭)만을 후속탐색에서 고려하고 나머지는 탐색으로부터 제외하는 방법이다. 정합은 입력 프레임과 식 (1)의 범위내의 노드에 대응하는 음향 모델과의 정합을 의미한다. 여기서 각 노드의 우도를 비교하여 상위 일정 개수를 선택한 후, 여기서부터 전개되어지는 노드들과 입력 $i+1$ 프레임과 정합한다.

탐색공간을 더욱 제한하는 방법으로써 프루닝 기법[17,18]이 있다. 이 방법은 각 프레임에 있어서 최대 우도를 $g_{\max}$ 로 하고, $g_{\max} - \lambda$ (λ는 여유분을 둔 문턱치)에 만족하지 않는 후보에 대해서는 그 시점 이후의 탐색을 프루닝함으로써 탐색공간을 감소시킨다.

먼저 One-pass Viterbi 알고리즘의 누적대수우도확률 $p_q^n(i,j)$의 i 프레임에 대한 최대치 $P_{\max}(i)$을 다음과 같이 구할 수 있다[19].

$$P_{\max}(i) = \max_{j,n,q} P_q^n(i,j) \tag{2}$$

이렇게 구해진 최대우도에 대해 식 (3)과 같은 조건을 만족하는 각 상태 q 의 각 단어 (또는 PLU; Phoneme Like Unit)에 대해서만 탐색을 수행하고 나머지는 제외하는 기법을 다음 식으로 나타낼 수 있다.

$$\max_j P_q^n(i,j) < P_{\max}(i) - \lambda \tag{3}$$

빔 탐색법에서 가장 중요한 것은 각 후보의 우도의 정도이다. 정도가 낮은 경우, 정해로 얻어진 후보가 프루닝에 의해 제외되는 오류가 있을 수 있다. 즉, 어떤 시점 (처리 프레임)에서 그 노드까지의 누적우도가 크지 않을 경우 정해가 될 수 있음에도 불구하고 탐색에서 제외되어 최적성을 보장받지 못하게 되므로, 빔 폭의 제한과 프루닝조건을 엄격하게 함으로써 최적해를 잃을 우려가 있다. 따라서, 인식정도에 영향을 주지 않기 위해서는 빔 폭과 프루닝조건을 완화시키면서 탐색공간을 감소시키는 방법을 찾을 필요가 있다.

2.3. 가변 프루닝 문턱치

2.2절에서 서술한 바와 같이 오토마타제어에 기초한 연속음성인식 알고리즘은 프레임에 동기하여 인식처리의 탐색공간을 확장하는 비교적 효율적인 방법이다. 그러나, 프레임의 수가 증가함에 따라 후속하는 오토마타의 상태수가 증가하게 되면, 계산량이 지수함수적으로 증가하게 되므로 대어휘 연속음성인식에 있어서는 치명적인 문제를 초래하게 된다. 이러한 계산량 증가의 문제를 해결하기 위해 2.1절에서의 목구조의 사전정보를 이용한 언어적 수법을 도입하고, 확률에 대한 제한을 통한 탐색공간의 감소법을 도입하면 대어휘에서도 효율적인 처리가 가능할 것으로 생각된다.

여기서는 프루닝 기법에 의한 탐색공간의 제한에 있어서 프레임의 진행에 따라 구해지는 누적확률에 근거하여 문턱치의 설정을 가변시키는 프레임동기형 가변 프루닝 문턱치의 도입을 제안한다. 이는 OPDP (One-Pass Dynamic Programming)법의 특성인 프레임 동기성의 장점을 이용하고, 최적 해를 위한 누적확률이 프레임의 진행에 따라 분포의 폭이 작아지며 수렴하는 것에 기초한 것이다. 전술한 식 (3)에 대하여 가변 프루닝 문턱치를 설정할 경우 다음과 같이 표현된다.

$$\max_{j} P_q^n(i,j) < P_{\max}(i) - \lambda(k) \qquad (4)$$

식 (4)에서 문턱치 λ 를 설정하는 데 있어서 프레임 i 에서의 정수값 k 에 따른 가변 프루닝 문턱치 $\lambda(k)$를 도입하고 있다. 정수값 k 의 설정은 프레임 i 에 대해 선형적 또는 비선형적으로 증가량을 설정할 수 있다.

누적대수우도확률과 $i \to j$ 프레임까지 진행되는 시점에서의 우도확률과의 차를 시간경과에 따라 나타낸 것으로 프레임 종단으로 갈수록 분포의 폭이 좁아짐을 알 수 있다. 이러한 가변 프루닝 문턱치를 사용함으로써 탐색의 공간을 더욱 제한함은 물론 2.2절에서 언급한 바와 같이 상수값으로 지정할 경우 일어날 수 있는 최적해를 잃지 않는 효과를 기대할 수 있다.

III. 프레임 단위 적응 프루닝 방법

2장에서 설명한 방법들이 각 프레임에서 후보 단어들을 이전에 제안된 방법들에 비해 보다 효과적으로 제한할 수 있었지만, 여전히 탐색할 필요가 없는 공간을 탐색한다. 따라서 여기서는 인식 과정 중에 탐색 공간을 효과적이고 자동으로 줄이기 위하여 프레임 단위 적응 프루닝 알고리즘을 제안한다.

이 알고리즘은 이웃 프레임사이의 최대 우도 확률들의 상관성이 크므로 앞 프레임의 최대 우도 확률로부터 효과적인 프루닝 문턱치를 얻을 수 있다는 점에 착안하여, 앞 프레임의 최대 우도 확률과 후보 우도 확률들의 조합으로 현재 프레임에서의 프루닝 문턱치를 프레임 단위로 갱신하는 방법이다.

현재 프레임의 프루닝 문턱치는 식 (5)를 이용하여 계산되어진다.

$$\lambda(k) = \frac{1}{N}\sum_{s=1}^{N}\{P_{\max}(i-1,j^*) - P_{hyp}(i-1,s)\} \qquad (5)$$

여기서, $P_{\max}(i-1,j^*)$는 프레임 $i-1$에서 최대 우도 확률이고, $P_{hyp}(i-1,s)$는 프레임 $i-1$에서 여러 후보들의 우도 확률이고, 그리고 N은 프레임 $i-1$에서 후보의 수이다.

식 (5)로부터 알 수 있는 바와 같이 제안된 알고리즘은 현재의 문턱치가 인식 과정 중에 얻어질 수 있기 때문에, 인식 태스크가 바뀌더라도 문턱치를 구하기 위하여 여러 번의 사전 실험을 필요로 하지 않는다. 또한, 문턱치가 적응적으로 얻어지기 때문에 다른 환경하에서도 인식 속도를 향상시킬 수 있다. 그림 2에 제안된 프레임단위 적응 프루닝 문턱치를 고정 프루닝과 가변 프루닝 문턱치와 비교하여 나타내었다. 그림에서 보는 바와 같이 적응 프루닝 문턱치를 사용하는 경우 문턱치가 커지는 경우가 발생하였는데, 이는 인식과정 중에 단어나 음소 사이에서 확률값이 갑자기 변하기 때문에 생기는 현상이다. 전체적인 탐색공간을 비교해 보면 고정 프루닝 문턱치와 가변 프루닝 문턱치에 비하여 탐색공간이 줄어드는 것을 알 수 있다.

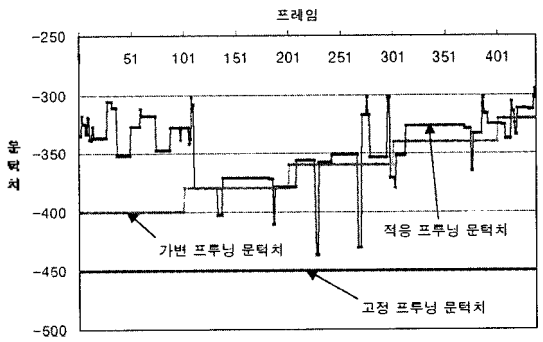


그림 2. 적응 프루닝 문턱치 (고정 프루닝 문턱치와 가변 프루닝 문턱치와 비교)

Fig. 2. An adaptive pruning threshold (comparing with fixed and variable pruning threshold).

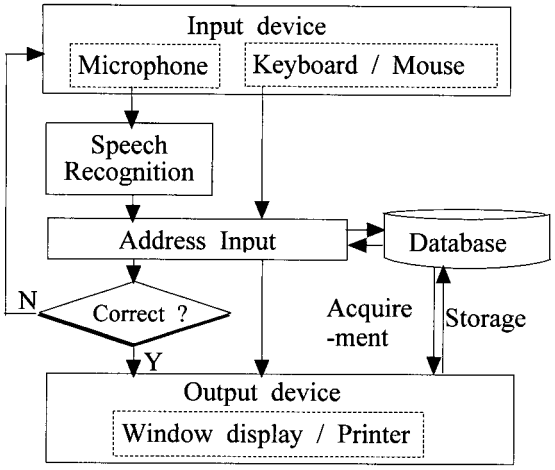


그림 3. 전체 시스템의 구성도
Fig. 3. Overview of the address recognition system.

IV. 인식 실험

4.1. 시스템 개요

제안한 적응 프루닝 문턱치 도입의 유효성을 확인하기 위하여 인식실험을 실시하여 인식률의 손상없이 탐색시간이 감소할 수 있는지를 확인한다. 인식시스템은 사운드카드를 내장한 개인용 컴퓨터 상에서 동작할 수 있고 국내의 행정 단위 주소단어를 인식의 대상으로 한다. 주소음성인식에 있어서는 광역 단위로부터 탑다운 방법으로 인식, 검색하게 된다. 이 시스템에의 입력은 음성과 컴퓨터의 기본 입력 장치인 키보드, 마우스 모두를 통해서도 가능하다. 주소입력에 있어서는 최초 광역 단위인 전국 시도의 후보 단어를 윈도우 화면상에 나타내고 인식결과에 따른 하위 행정 단위의 후보만을 화면에 보여줌으로써 사용자가 주소를 선택하는 데 있어서의 오류를 방지할 수 있도록 하고 있다. 그림 3에 시스템의 전체 구성도를 나타낸다.

4.2. 음성데이터 및 분석

음성 데이터로는 화자독립 기본모델 (SI-HMM; Speaker Independent HMM)의 작성을 위하여 한국전자통신연구원 (ETRI)에서 작성한 PBW (Phoneme Balanced Words) 445단어 음성 데이터베이스 중 14인의 1회 발성을 이용한다. 적응화 단계에 있어서는 사무실 환경에서 3인의 남성 화자가 데스크탑 마이크를 이용하여 발성한 100개의 연결주소단어 중 25개의 연결주소단어를 이용하여 SI-HMM을 적응화한다. 인식단계에서는 나머지 75개의 연결주소단어를 사용한다. 상위 클래스에서 하위 클래스로 단계적으로 인식이 진행되는 한국어 주소의 특징을 고려한 유한상태 오토마타를 구성하였고, 전국의 모든 행정 단위를 포함하고 있다. 표 1에 음성데이터와 시스템 환경을 나타낸다.

표 2. 음성 데이터의 분석조건
Table 2. Analysis condition of speech data.

Sampling Frequency	16 kHz
Resolution	16 bits
Hamming Window	16 msec (256 points)
Frame Rate	5 msec (80 points)
Analysis	14 order LPC
Feature Parameters	10 order MFCC+10 order RGC

발성된 주소음성 데이터는 16 kHz/16 bits로 A/D변환된 후, 에너지와 영교차율의 평균과 분산을 이용하여 음성 구간을 검출한다. 검출된 주소음성으로부터 14차의 LPC분석을 통하여 10차의 멜켑스트럼 계수 (MFCC; Mel Frequency Cepstral Coefficient)를 구하고, 이 멜켑스트럼 계수와 그 회귀계수 (RGC; Regressive Coefficient)를 음성특징 파라미터로 한다. 인식실험시 10차의 MFCC와 10차의 RGC를 사용한다. 표 2에 특징 추출을 위한 음성자료의 분석 조건을 나타낸다.

4.3. HMM 학습과 적응화

주소음성인식을 위한 인식의 기본단위는 48개의 유사음소로 하고, 각 유사음소의 음향모델에 사용한 HMM은 4상태 3출력분포의 연속출력분포형 HMM (이산분포형 지속시간 제어)을 사용한다. 마이크의 변동, 사용환경변화에 대한 인식률 제고를 위하여 최대사후확률분포를 이

표 1. 학습, 적응화, 인식용 음성데이터
Table 1. Speech data for training, adaptation and recognition.

Speaker (Number)	Male (14)	Male (3)	Male (3)
Utterance Type (#)	PBWs (445)	Connected Words (25)	Connected Words (75)
# of utterance	1	1	1
Usage	Training	Adaptation	Recognition
Environment	Soundproof Booth	Office	Office
Record Device	DAT Recorder	PC (Sound Card)	PC (Sound Card)
Microphones	Dynamic Headset	Condenser Desktop	Condenser Desktop

용한 적응화 기법[20]을 이용하여 음소 HMM모델 (SI-HMM)을 적응화하고, 인식단계에서는 적응화된 음소모델을 이용하여 OPDP알고리즘으로 인식한다. 적응화와 인식은 모두 사무실 환경하에서 이루어진다.

일반적으로 학습용 데이터와 평가용 데이터의 사이에 마이크를 포함한 녹음 환경의 차이, 화자 변화에 따른 발성의 차이 등은 인식률에 직접적인 영향을 미친다. 이를 해소하기 위해 적응화 학습을 실시한다. 기존의 HMM을 이용한 음성인식기의 대부분은 ML (*Maximum Likelihood*)에 기반을 둔 Baum-Welch 학습법으로 파라미터를 재추정하고 있다. ML학습은 기본적으로 무한한 양의 학습데이터와 각 모델이 서로 독립적이라는 가정을 기초로 한다. 그러나 실제적인 학습의 경우 각 모델들은 서로 독립적으로 보기 어렵고 학습데이터 양도 상당히 제약되어 있어서 인식기의 변별력을 저하시키는 원인이 된다. 반면에 최대사후확률추정법을 이용하면 적응화가 중단되어도 그 시점까지 최적인 파라미터를 추정할 수 있고 필요시 추가적으로 적응화를 수행해 파라미터의 정밀도를 향상시킬 수 있다.

4.4. 연결단어 인식

주소인식 시스템에서 인식의 대상이 되는 주소의 경우 고립단어가 여러 개 나열된 연결단어 형태로 구성되어 있기 때문에 고립단어를 대상으로 하는 인식방법으로는 제약이 따르게 된다. 만약 고립단어인식법으로 연결 주소명을 인식하는 경우, 한 단어를 발성하고 이를 인식하여 결과를 확인하고 또 다시 하위 단위의 단어를 발성하여 인식을 수행해야 하기 때문에 사용상 불편이 크다.

따라서 주소의 경우 연결단어 즉, "대구광역시 수성구 만촌동" 등과 같은 주소명 특유의 연속적인 발성으로 확장하여야 한다. 이를 위하여 본 논문에서는 주소인식 시스템의 인식대상이 주소라는 점을 고려한 연결단어를 위한 유한상태 오토마타 (Finite State Automata)를 구성한다. 고립단어인식에서 연결단어로 인식 대상을 확장하기 위한 전국 주소명에 대한 유한상태 오토마타를 구성한다. 처음 시작상태에서는 전국 광역시도 단위를 대상으로 인식을 수행하고, 인식된 시도 단위에 해당하는 하위 행정단위를 인식하며 같은 방법으로 그 하위 행정 단위를 인식하도록 계층적인 구조를 이용한다.

V. 인식 실험 결과

제안된 방법의 유효성을 확인하기 위하여 기존의 방법인 고정 프루닝 문턱치, 가변 프루닝 문턱치, 그리고 제안된 적응 프루닝 문턱치 알고리즘을 적용하여 인식 실험을 수행한 후, 세 가지 알고리즘을 비교하였다. 인식 실험은 3인의 화자가 발성한 75개의 연결단어에 대하여 탐색 공간을 측정하였고, 워크스테이션 (167 MHz)상에서 off-line으로 수행되었다. 주소에 대하여 인식률 (CWRR; Connected Word Recognition Rate)과 탐색 공간 (SS; Search Space)을 측정하였고, 주소를 이루는 각 단어에 대하여 단어 인식률 (WRR; Word Recognition Rate)을 3인의 화자에 대하여 구하였다. 탐색 공간을 측정한 이유는 일반적으로 컴퓨터 성능에 따라 같은 인식 태스크라 하더라도 인식 시간이 달라질 수 있다. 따라서 신뢰성 있는 결과를 얻기 위하여 탐색되는 전체 단어와 후보 단어를 고려하여 식 (6)를 이용하여 탐색 공간을 측정하였다.

$$SS(\%) = \frac{N_{match}}{N_{match} + N_{skip}} \times 100.0 \qquad (6)$$

여기서, N_{match}는 각 프레임에서 탐색되는 단어의 평균수이고, N_{skip}은 각 프레임에서 탐색되지 않는 단어의 평균수를 나타낸다.

제안된 방법의 유효성을 확인하기 위해, 앞에서 설명한 고정 프루닝, 가변 프루닝, 적응 프루닝 알고리즘을 적용하여 인식실험을 수행한 후 비교하였다. 앞에서 설명한 고속화 기법 중 고정 프루닝 문턱치를 이용하여 인식실험을 실시하였다. 이때 HMM 모델은 적응화 후의 모델을 사용하고, 빔 폭은 10으로 하여 프루닝 문턱치의 값만

표 3. 고정 프루닝 문턱치를 이용한 인식실험 결과
Table 3. Recognition results using fixed pruning threshold.

Pruning threshold	CWRR (%)	WRR (%)	SS (%)
-500	96.0	98.7	30.96
-400	96.0	98.7	30.63
-300	95.7	98.5	30.46

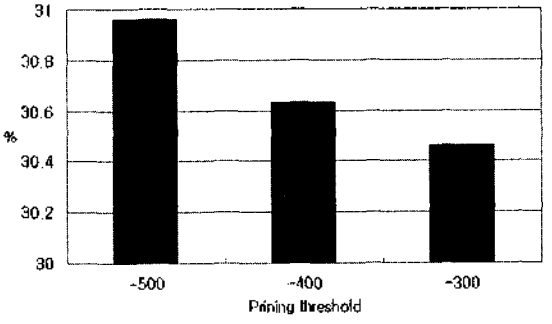


그림 4. 고정 프루닝 문턱치의 탐색공간 비교.
Fig. 4. Search space using fixed pruning threshold.

표 4. 가변 프루닝 문턱치를 이용한 인식실험 결과
Table 4. Recognition results using variable pruning threshold.

Pruning threshold	CWRR (%)	WRR (%)	SS (%)
FSV1	96.0	98.7	29.17
FSV2	96.0	98.7	29.11
FSV3	96.0	98.7	29.09
FSV4	96.0	98.7	28.99
FSV5	96.0	98.7	28.87
FSV6	95.7	98.5	28.80

단, FSV1: -400, -370, -350, -320, -290, -260
FSV2: -400, -380, -360, -330, -300, -230
FSV3: -400, -380, -360, -300, -300, -190
FSV4: -360, -340, -320, -300, -300, -10
FSV5: -310, -310, -300, -290, -270, -10
FSV6: -310, -310, -300, -290, -270, -10

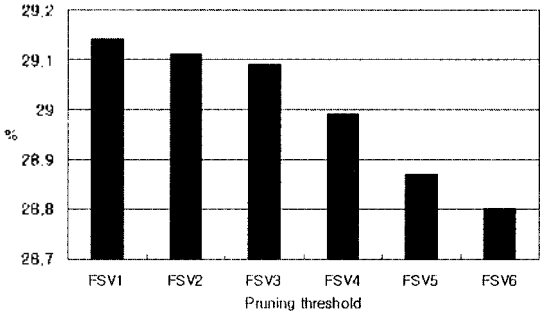


그림 5. 가변 프루닝 문턱치의 탐색공간 비교
Fig. 5. Search space using variable pruning threshold.

표 5. 적응 프루닝 문턱치를 적용한 인식 실험 결과
Table 5. Recognition results by an adaptive pruning threshold.

Pruning threshold	CWRR (%)	WRR (%)	SS (%)
Adaptive	96.0	98.7	26.23

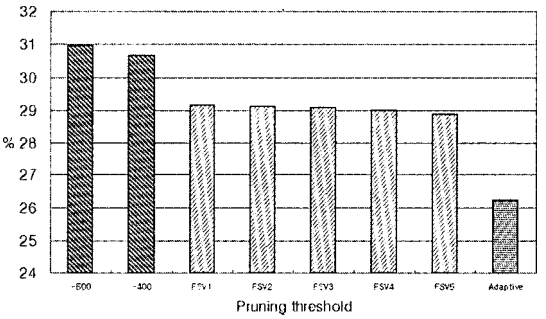


그림 6. 고정 프루닝, 가변 프루닝, 적응 프루닝 문턱치의 탐색공간 비교
Fig. 6. Search space compared with fixed pruning, variable pruning and adaptive pruning threshold.

을 변화시키면서 인식실험을 수행하였다. 이 결과를 표 3과 그림 4에 나타내었다.

프루닝 문턱치가 400인 경우, 인식률의 저하없이 탐색공간이 효과적으로 줄어드는 것을 알 수 있다. 실제로 개인용 컴퓨터 환경하에서 동작시켰을 경우 별도의 하드웨어 추가 없이 연결단어의 평균 발성시간이 약 2.5초 정도가 소요될 경우, 발성이 끝나는 시점으로부터 약 2초 내에 인식이 완료되어 실시간 처리가 가능함도 알 수 있었다[10].

다음은 빔 탐색법의 문제점을 고려하여 프레임이 진행됨에 따라 가변되는 프레임동기형 프루닝 문턱치를 이용하여 인식실험을 수행하였다. 이때의 인식실험 결과를 표 4와 그림 5에 나타내었다.

가변 프루닝 문턱치를 사용하였을 경우 표에서 보는 바와 같이 인식률이 저하되지 않으면서 기존의 고정 프루닝 문턱치보다는 1.76%의 탐색공간이 감소되어 가변 프루닝 문턱치 알고리즘의 유효성을 확인할 수 있었다.

다음은 적응 프루닝 문턱치의 인식 실험 결과를 표 5에 보였고, 제안된 방법을 기존의 방법과 비교해서 그림 6에 나타내었다. 신뢰성 있는 탐색 공간을 알기 위하여 인식시간이 아닌 매칭되는 후보수로서 탐색 공간으로 구하였다. 그림에서 보는바와 같이 적응 프루닝 문턱치는 기존의 고정 프루닝 문턱치와 가변 프루닝 문턱치에 비해 탐색공간이 각각 4.4%와 2.64%의 탐색공간이 감소되어 제안된 방법의 유효성을 확인할 수 있었다.

VI. 결 론

본 논문에서는 불필요한 탐색 공간을 보다 효과적으로 제한하기 위하여 인식 과정 중에 탐색공간을 효과적이면서 자동적으로 제한하기 위한 프레임 단위 적응 프루닝 문턱치 알고리즘을 제안하여, 전국의 주소를 대상 어휘로 하는 한국어 주소인식 시스템에 적용하여 제안한 고속화 알고리즘의 유효성을 확인하였다.

프레임 단위 적응 프루닝 알고리즘은 이웃 프레임사이의 최대 확률의 상관성이 큰 점에 착안하여, 앞 프레임의 최대 확률로부터 효과적으로 프루닝 문턱치를 얻는 방법으로 현재 프레임에서 적응 프루닝 문턱치는 앞 프레임의 최대 확률과 후보 확률의 조합으로 결정할 수 있다.

제안된 방법의 유효성을 확인하기 위하여 한국어 주소인식 시스템에 적용한 후 인식실험을 수행한 결과, 연결단어 인식률이 96.0%와 단어 인식률이 98.7%인 경우를 기준으로 하였을 때 제안된 알고리즘이 고정 프루닝과 가변 프루닝 문턱치에 비하여 인식률의 저하없이 14.4%와 9.14%의 탐색 공간을 상대적으로 줄일 수 있음을 확인할 수 있어 제안된 방법의 유효성을 확인할 수 있었다.

감사의 글

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

참 고 문 헌

1. 정현열, "음성인식 연구의 국내외 현황과 전망," 제15회 음성통신 및 신호처리 워크샵 논문집, pp. 23-30, 8, 1998.
2. 김순협, "음성인식의 현황과 최근 연구 동향," 2000년도 한국음향학회 학술발표대회 논문집, Vol. 19, No. 2(s), 11, 2000.
3. M. K. Ravishankar, "Efficient Algorithms for Speech Recognition," Ph. D Thesis, Carnegie Mellon University, 1996.
4. Alleva, F., et al, "Applying SPHINX-II to the DARPA Wall Street Journal CSR task," *Proc. of Speech and Natural Language Workshop*, pp. 393-398, Feb. 1992.
5. Alon Lavie, et al, "JANUS-III: Speech-to-speech translation in multiple languages," *Proc. IEEE ICASSP-97*, Vol. 1, pp. 99-102, April 1997.
6. A. Kai and S. Nakagawa, "A frame-synchronous continuous speech recognition algorithm using a top-down parsing of context-free grammar," *Proc. ICSLP 92*, pp. 257-260, 1992.
7. T. Nishimoto, N. Shida, T. Kobayashi, K. Shirai, "Multimodal Drawing Tool Using Speech, Mouse and Keyboard," *Porc. ICSLP*, Vol. 3, pp. 1287-1290, 1994.
8. Katsuhiko Shirai, "Spoken Dialogue in Multimodal Human Interface," *ICSP '97*, pp. 13-20, Aug. 1997.
9. Hyun-Yeol Chung, Cheol-Jun Hwang and Shi-Wook Lee, "A Bimodal Korean Address Entry/Retrieval System," *ICSLP98*, pp. 1607-1610, 12, 1998.
10. 김득수, 황철준, 정현열, "가능을 가진 주소입력 시스템의 개발과 평가," 한국음향학회지 제18권 제2호, pp. 3-10, 2, 1999.
11. 황철준, 오세진, 김범국, 정호열, 정현열, "실시간 주소인식을 위한 시스템의 인식속도 개선," 1999년도 한국음향학회 하계학술대회 논문집, 제18권 제1(s)호, pp. 74-77, 7, 1999.
12. Cheol-Jun Hwang, Se-Jin OH, Ho-Youl Jung and Hyun-Yeol Chung, "An Adaptive Pruning Threshold Algorithm for Efficient Speech Recognition," *SPECOM '99*, pp. 103-106, 10, 1999.
13. 황철준, 오세진, 김범국, 정호열, 정현열, "음성인식의 고속화를 위한 프레임단위 프루닝 알고리즘," 2000년도 한국음향학회 정기총회 및 학술발표대회 논문집, 제19권 제2(s)호, pp. 183-186, 11, 2000.
14. P. S. Gopalakrishnan, L. R. Bahl and R. L. Mercer, "A tree search strategy for large-vocabulary continuous speech recognition," *Proc. IEEE ICASSP-95*, Vol. 1, pp. 572-575, May 1995.
15. F. Richardson, M. Ostendorf and J. R. Rohlicek, "Lattice-based search strategies for large vocabulary speech recognition," *Proc. IEEE ICASSP-95*, Vol. 1, pp. 576-579, May 1995.
16. S. Furui, "Speaker-Independent Isolated Word Recognition Using Dynamic Features of Speech Spectrum," *IEEE Trans. Acoustics, Speech and Signal Processing*, Vol. 34, No. 1, pp. 52-59, Feb. 1986.
17. 坂井利, 中川聖一, "構文情報を用いた連續音聲認識," 情報處理學會 第15回 全國大會, pp. 37, Dec. 1994.
18. B. T. Lowerre, "HARPY Speech Recognition System," Ph. D thesis, Carnegie Mellon University, 1976.
19. J. C. Junqua, and J. P. Haton, "Robustness in Automatic Speech Recognition," Kluwer Academic Publishers, 1996.
20. L. Rabiner and B. H. Juang, "Fundamentals of Speech Recognition," Prentice-Hall International, Inc, 1993.

저자 약력

● 황 철 준 (Chul-Jun Hwang)

1996년 2월: 영남대학교 전자공학과 (공학사)
1998년 2월: 영남대학교 대학원 전자공학과 (공학석사)
1998년 3월~현재: 영남대학교 대학원 전자공학과 (박사수료)
2000년 3월~현재: 대구과학대학 정보전자통신계열 전임강사
※ 주관심분야: 음성분석 및 인식, 디지털 신호처리

● 오 세 진 (Se-Jin Oh)

1996년 2월: 영남대학교 전자공학과 (공학사)
1998년 2월: 영남대학교 대학원 전자공학과 (공학석사)
1998년 3월~현재: 영남대학교 대학원 전자공학과 (박사수료)
※ 주관심분야: 음성분석 및 인식, 언어처리

● 김 범 국 (Bum-Koog Kim)

1990년 2월: 영남대학교 수학과 (이학사)
1992년 2월: 영남대학교 대학원 전자공학과 (공학석사)
1998년 2월: 영남대학교 대학원 전자공학과 (공학박사)
1997년 3월~현재: 대구과학대학 정보전자통신계열 조교수
※ 주관심분야: 음성분석 및 인식, 언어처리, 멀티모달 시스템

● 정 호 열 (Ho-Youl Jung)

1988년 2월: 아주대학교 전자공학과 (공학사)
1990년 2월: 아주대학교 전자공학과 (공학석사)
1993년 2월: 아주대학교 전자공학과 (박사수료)
1998년: (프)리옹국립응용과학원 (INSA de Lyon) 전자공학전공 (공학박사)
1998년 4월~1998년 12월: (프)CREATIS 박사후 과정
1999년 3월~현재: 영남대학교 전자정보공학부 조교수
※ 주관심분야: 음성·영상 신호처리, 인공지능, 디지털 워터마킹 등

● 정 현 열 (Hyun-Yeol Chung)

현재: 영남대학교 전자정보공학부 교수
한국음향학회지 제19권 제6호 참조