

# 실시간 주소 음성인식을 위한 인식 시스템의 인식속도 개선

황철준\*, 오세진\*, 김범국\*\*, 정호열\* 정현열\*

\* 영남대학교 정보통신공학과

\*\* 대구과학대학 전자과

## Improvement of Recognition Speed for Real-time Address Speech Recognition

Cheol-Jun Hwang\*, Se-Jin Oh\*, Bum-Koog Kim\*\*, Ho-Youl Jung\*, Hyun-Yeol Chung\*

\* Department of Information and Communication Eng., Yeungnam University

\*\* Department of Electronics, Taegu Science College

### 요 약

### I. 서 론

본 논문에서는 본 연구실에서 개발한 주소 음성인식 시스템의 인식 속도를 개선시키기 위하여 새로운 가변 프루닝 문턱치를 적용하는 방법을 제안하고 실험을 통하여 그 유효성을 확인하였다.

기존의 가변 프루닝 문턱치는 일정 프레임이 경과하면 일정 값을 가진 문턱치를 계속하여 감소시켜나가는 방법을 반복하기 때문에, 불필요한 탐색공간을 탐색하게 된다. 본 논문에서 새로이 제안하는 가변 프루닝 문턱치를 채용하는 방법은 처음 일정 구간이 경과되면 일정 문턱치를 감소시키거나, 다음 일정 프레임에서는 탐색되어야 할 후보에 따라서 문턱치를 변화시켜 프루닝시키기 때문에 탐색공간을 효과적으로 감소시킬 수 있다.

제안된 방법의 유효성을 확인하기 위하여, 본 연구실에서 개발한 한국어 주소 입력 시스템에 적용하였다. 이 시스템은 48개의 연속 HMM 유사음소단위(Phoneme Like Units; PLUs)를 인식의 기본단위로 하고, 사용환경 변화에 의한 인식성능의 저하를 최소화하기 위해 최대사후 확률추정법(Maximum A Posteriori Probability Estimation; MAP)을 사용하며, 인식알고리즘으로는 OPDP(One Pass Dynamic Programming)법을 이용하고 있다.

남성화자 3인에 의한 75개의 연결주소명을 이용하여 인식 실험을 수행한 결과 고정 프루닝 문턱치를 적용한 경우 인식률은 평균 96.0%, 인식 시간은 5.26초였고, 기존의 가변 프루닝 문턱치의 경우 인식률은 평균 96.0%, 인식 시간은 5.1초인 데 비하여, 새로운 가변 프루닝 문턱치를 적용한 경우에는 인식률 저하없이 인식 시간이 4.34초로, 기존에 비해 각각 0.92초, 0.76초 인식 시간이 감소되어 제안한 방법의 유효성을 확인할 수 있었다.

최근 정보 통신 분야의 비약적인 발전과 더불어 개인용 컴퓨터 보급이 가속화됨에 따라 인간과 기계사이의 대화에 대한 연구의 중요성이 증대되고 있다.

음성인식에 관한 연구는 최근 몇 년간의 눈부신 발전으로 일부 고립단어를 인식할 수 있는 상업용 시스템이 구현되었다. 이러한 실용화를 위한 시스템에서 중요한 것은 고정도의 인식률과 함께 실시간 인식이 필수적이다. 고립단어, 또는 한정된 태스크 범주의 연속음성인식의 경우 잠음환경하에서도 95%이상의 인식률을 가진 시스템이 많이 개발되고 있다.

인식시간의 경우, 외국에서는 고립단어 뿐만 아니라 한정된 태스크의 연속음성 인식에서도 거의 실시간으로 동작하는 시스템이 많이 개발되고 있으나, 국내의 경우 고립단어 인식에 있어서는 실시간 동작이 가능하지만, 연속음성 인식의 경우에는 아직까지 많은 연구가 요구되고 있는 실정이다.

본 연구실에서는 지금까지 대어휘 연속음성인식 시스템의 인식시간을 실시간 처리하기 위한 연구를 수행하여 왔다. 특히, 주소 입력 시스템의 경우, 실시간 처리의 중요성이 강조되어 현재까지 고정 프루닝 문턱치를 이용하여 프레임 구간이 진행됨에 따라 프루닝 문턱치로서 하나의 일정한 값을 사용하여 탐색 공간을 감소시키는 방법[1], 이를 개선한 가변 프루닝 문턱치를 이용한 경우에는 프레임 구간이 진행함에 따라 일정한 비율로 문턱치를 감소시켜 불필요한 탐색 공간을 줄이고 있다 [2]. 이 방법들은 인식 시간을 효과적으로 줄일 수 있으나 여전히 탐색할 필요가 없는 후보에 대해서도 탐색작업을 수행하게 됨으로서 많은 인식시간을 필요로 하는

문제점을 안고 있다.

이러한 문제점을 해결하기 위하여, 본 연구에서는 탐색 공간을 더욱 효과적으로 줄이기 위해 새로운 가변 프루닝 문턱치를 제안하고 이를 실험을 통해 그 유효성을 확인하고자 한다.

2절에서는 기존에 사용한 프루닝 문턱치 설정 방법에 대하여 설명하고, 3절에서 새로운 가변 프루닝 문턱치에 대하여 소개한다. 4절에서는 실험에 이용한 시스템 개요와 음성 데이터를, 5절에서 인식실험 및 결과에 대하여 설명하고, 마지막으로 6장에서 결론을 맺는다.

## II. 기존의 프루닝 문턱치

일반적으로 음성인식은 예측된 전체 후보와 입력음성을 정합시키는 방법을 이용하는 데, 대상 어휘수가 증가하고 인식 알고리즘이 복잡해짐에 따라 대규모 탐색공간이 필요하며 이에 따라 많은 처리시간을 요구된다. 따라서 실시간 음성인식을 위해서는 전체의 후보와 정합을 행하지 않고도 고정도의 인식성능을 얻을 수 있는 효과적인 탐색수법이 필요하다.

### 2.1 고정 프루닝 문턱치

주어진 입력음성에 대하여 우도가 가장 큰 후보를 탐색하는 확률적 음성인식의 방법에 있어서, 모든 가능한 후보를 고려할 경우 어휘수의 증가에 따라 탐색공간이 지수함수적으로 증가하게 되므로, 많은 계산량과 메모리가 필요하게 되어 비현실적이다.

한편, 빔 탐색법(Beam Search)[3][4]은 각 프레임에서 몇 개의 부분경로만을 추정하기 때문에 인식대상 어휘수의 증가와 관계없이 일정한 수준이하로 탐색 공간을 줄일 수 있다. 이 방법은 각 후보의 우도를 비교하여 상위 일정 개수(문턱치 이하의 것)에 대해서만 후속의 정합을 고려하는 방법으로 다음과 같이 나타낸다.

$$\max_j P_q^n(i, j) \leq P_{\max}(i) - \lambda \quad (1)$$

이 방법에 의하면  $i$  프레임에서의 최적 경로에 대해 문턱치  $\lambda$  이내의 상위 몇 개(빔 폭)만을 후속의 탐색에서 고려하고 나머지는 프루닝된다[5]. 그러나, 이 방법은 전 프레임에 대하여 일정한 문턱치값을 적용하기 때문에 불필요한 후보를 탐색하게 되는 문제점이 있다.

### 2.2 가변 프루닝 문턱치

고정 프루닝 문턱치의 문제점을 해결하기 위하여 프레임이 진행함에 따라 탐색공간을 유동적으로 제한하기 위하여 많은 연구에서 가변 프루닝 문턱치를 도입하고 있다[6]. 이 방법은 OPDP의 특징인 프레임 동기성을 이

용하며 최적해가 누적 확률이 프레임의 진행에 따라 분포의 폭이 작아지며 수렴한다는 것에 근거한 것이다. 식 (1)에 대하여 가변 프루닝 문턱치를 설정하는 경우 다음과 같이 표현된다.

$$\max_j P_q^n(i, j) \leq P_{\max}(i) - \lambda(k) \quad (2)$$

위 식에서 문턱치  $\lambda$ 를 설정할 때 프레임  $i$ 에서의 정수값  $k$ 에 따른 가변 프루닝 문턱치  $\lambda(k)$ 를 사용한다.

## III. 제안된 가변 프루닝 문턱치

기존의 가변 프루닝 문턱치의 경우에는 고정 프루닝 문턱치를 보완하기 위하여 프레임이 진행함에 따라 일정한 비율로 문턱치를 줄여가면서 선형적으로 탐색 공간을 감소시키고 있다. 이 방법은 인식 시간을 효과적으로 줄일 수 있지만, 여전히 탐색의 필요성이 없는 후보에 대해서도 탐색을 수행하게 된다.

이러한 문제점을 해결하기 위하여, 본 연구에서는 탐색이 진행됨에 따라서 탐색 공간을 더욱 엄격하게 줄일 수 있도록 새로운 가변 프루닝 문턱치를 적용하는 방법을 제안한다.

$$\max_j P_q^n(i, j) \leq P_{\max}(i) - \lambda(k^*) \quad (3)$$

여기서  $\lambda(k^*)$ 는 프레임  $i$ 에서 후보의 확률값에 따라서 변화시키게 된다.

제안한 새로운 가변 프루닝 문턱치는 일정 프레임이 지나면 일정 문턱치를 감소시킨 후 다음 일정 프레임에서는 실제 탐색 공간에 따라서 문턱치를 변화시켜 주면서 프루닝시킨다. 따라서 효과적으로 탐색공간을 줄일 수가 있다.

그림 1에 새로운 가변 프루닝 문턱치를 적용하는 방법을 고정 프루닝 문턱치, 기존의 가변 프루닝 문턱치를 적용하는 방법과 비교하여 나타낸다.

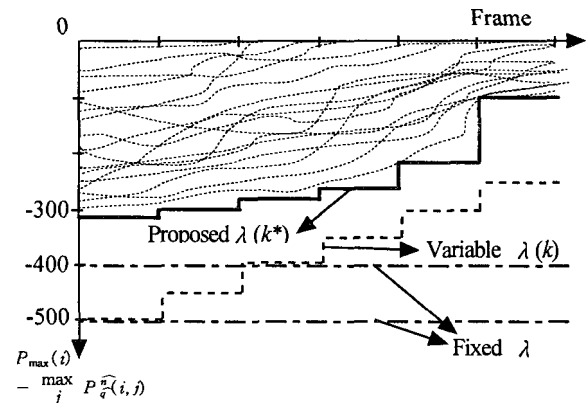


그림 1. 새로운 가변 프루닝 문턱치.

#### IV. 인식 시스템 개요와 음성 데이터

제안된 방법의 유효성을 확인하기 위하여, 본 연구실에서 개발한 한국어 주소 입력 시스템에 적용하였다. 본 시스템은 48개의 연속 HMM 유사음소단위(Phoneme Like Units; PLUs)를 인식의 기본단위로 하고, 사용환경 변화에 의한 인식성능의 저하를 최소화하기 위해 최대사후확률추정법(Maximum A Posteriori Probability Estimation; MAP)을 사용하고, 인식알고리즘으로는 OPDP(One Pass Dynamic Programming)법을 이용한다. 시스템의 학습과 적응화, 그리고 인식과정의 전체 흐름도를 그림 2에 나타내었다.

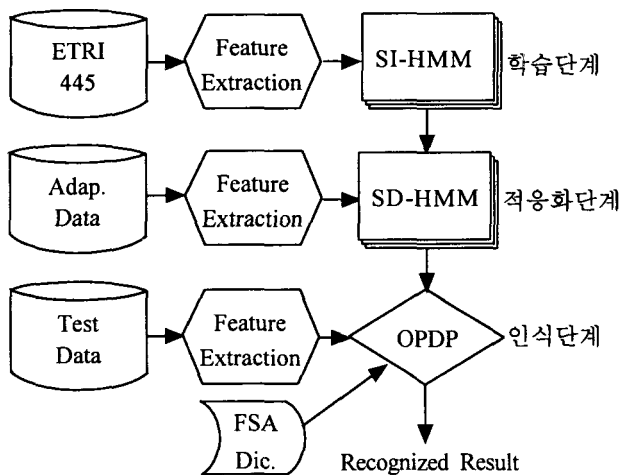


그림 2. 학습 및 평가의 전체 흐름도.

음성 자료로는 화자독립 기본모델(Speaker Independent HMM; SI-HMM)의 작성을 위하여 한국전자통신연구원(ETRI)에서 작성한 PBW(Phoneme Balanced Words) 445단어 음성 데이터베이스중 14인의 1회 발성을 이용한다. 적응화단계에 있어서는 사무실환경에서 3인의 남성화자가 데스크탑 마이크를 이용하여 발성한 100개의 연결주소단어중 25개의 연결주소단어를 이용하여 SI-HMM을 적응화한다. 인식단계에서는 나머지 75개의 연결주소단어를 사용한다. 유한상태 오토마타는 전국의 모든 행정단위를 포함하고 있다.

표 1에 실험에 사용된 음성 데이터베이스를 나타내었다.

표 1. 음성 데이터베이스.

화자	발성회수	발성어휘수	발성환경	사용
14인	1회	445	방음부스	학습
3인	1회	25	사무실	적응화
3인	3회	75	사무실	인식

#### V. 인식 실험 및 결과

인식실험을 위해서는 앞 절에서 설명한 시스템을 사용하였다. 실험에 사용한 음성데이터는 사무실 환경에서 100개의 주소명을 남성화자 3인이 콘텐서 데스크탑 마이크로 발성한 연결주소 데이터를 사용하였다.

실제 음성인식 시스템에 사용되는 여러 환경에 대하여 고정도 인식률을 얻기 위해서 최대사후확률추정법에 의한 적응화 기법을 도입하여 초기 화자독립 HMM을 적응화하였다. 적응화를 위해서는 대구 광역시의 행정단위를 포함하는 25개를 이용하였다. 인실 실험에서 빔 폭은 기존의 연구결과로부터 10을 사용하였다[7].

기존의 고정 프루닝 문턱치와 가변 프루닝 문턱치를 사용한 경우의 인식 실험 결과를 표 2에 나타내었다.

표 2. 고정 프루닝 문턱치와 가변 프루닝 문턱치에 의한 인식 성능과 인식 시간.

Fixed pruning threshold		
Pruning threshold	CWRR(%)	Time(sec)
-500	96.0	6.05
-400	96.0	5.26
-300	95.6	4.24
-200	94.7	3.03
Variable pruning threshold		
Pruning threshold	CWRR(%)	Time(sec)
FSV1	96.0	5.10
FSV2	95.1	4.78
FSV3	92.9	4.06
FSV4	95.6	4.47
FSV5	88.9	3.35

단, FSV1: 초기치=-500, 100프레임마다 50씩 감소  
 FSV2: 초기치=-500, 100프레임마다 60씩 감소  
 FSV3: 초기치=-500, 50프레임마다 50씩 감소  
 FSV4: 초기치=-400, 100프레임마다 50씩 감소  
 FSV5: 초기치=-400, 50프레임마다 50씩 감소

이 결과에서 고정 프루닝 문턱치의 경우 인식률 96.0%, 인식 시간은 5.26초였고, 기존의 가변 프루닝 문턱치의 경우는 인식률 저하없이 인식 시간이 5.10초로 약 0.16초 정도 인식 속도가 개선되었다.

본 논문에서 제안된 새로운 가변 프루닝 문턱치를 적용한 경우의 인식 실험 결과 인식률과 인식 시간을 함께 나타내었다.

표 3. 새로운 가변 프루닝 문턱치를 적용한 인식 실험 결과.

Pruning threshold	CWRR(%)	Time(sec)
NFSV1	96.0	4.79
NFSV2	96.0	4.68
NFSV3	96.0	4.51
NFSV4	96.0	4.42
NFSV5	96.0	4.34
NFSV6	95.6	4.30

단, 다음의 값은 100프레임마다 변화하는 값임.

NFSV1 : -400, -380, -360, -330, -300, -260  
 NFSV2 : -400, -380, -360, -330, -300, -230  
 NFSV3 : -400, -380, -360, -330, -300, -190  
 NFSV4 : -360, -340, -320, -300, -300, -10  
 NFSV5 : -340, -320, -300, -300, -300, -10  
 NFSV6 : -310, -310, -300, -290, -280, -10

새로운 가변 프루닝 문턱치를 적용한 경우 인식률의 저하없이 기존의 고정 프루닝 문턱치 보다는 0.92초, 기존의 가변 프루닝 문턱치보다는 0.76초의 인식 시간이 감소되어 제안한 방법의 유효성을 확인할 수 있었다.

## VI. 결론

본 논문에서는 기존에 본 연구실에서 개발한 주소 음성인식 시스템의 인식 속도를 개선시키기 위하여 새로운 가변 프루닝 문턱치를 적용하는 방법을 제안하였다.

기존의 가변 프루닝 문턱치는 일정 프레임이 경과할 때 일정값의 문턱치를 줄여나가는 과정을 반복하면서 불필요한 탐색공간을 감소시키고 있지만 여전히 불필요한 탐색공간을 탐색하게 된다. 제안한 새로운 가변 프루닝 문턱치는 처음 일정 프레임이 경과할 때까지 일정 문턱치를 줄이지만, 다음 일정 프레임에서는 탐색되어야 할 후보에 따라서 문턱치 값을 변화시켜 프루닝시키기 때문에 효과적으로 탐색공간을 줄일 수가 있었다.

제안된 방법의 유효성을 확인하기 위하여, 본 연구실에서 개발한 한국어 주소 입력 시스템에 적용하였다. 남성화자 3인이 75개의 연결주소명으로 인식 실험을 수행한 결과 고정 프루닝 문턱치를 적용한 경우 인식률은 평균 96.0%, 인식 시간은 5.26초였고, 기존의 가변 프루닝 문턱치의 경우 인식률은 평균 96.0%, 인식 시간은 5.1초였는데, 새로운 가변 프루닝 문턱치를 적용한 경우에는 인식률 저하없이 인식 시간이 4.34초로, 기존에 비해 각각 0.92초, 0.76초 인식 시간이 감소되어 제안한 방법의 유효성을 확인할 수 있었다.

향후 연구 진행 방향으로는 현재 방법은 여러 번의 실험을 거쳐 문턱치 값을 찾아내기 때문에 선행 실험이 많이 수행되어야 한다. 따라서, 문턱치 값을 자동적으로 적

용하는 방법에 대하여 연구를 수행할 예정이다.

## 참고 문헌

1. 이시욱, "주소입력시스템을 위한 음성인식의 고속화에 관한 연구," 영남대 석사학위 논문, 1997.6.
2. 김득수, 황철준, 정현열, "음성인식 기능을 가진 주소 입력 시스템의 개발과 평가," 한국음향학회지 Vol.18, No.2, 1999.2.
3. 坂井利, 中川聖一 "構文情報を用いた連続音聲認識," 情報處理學會 第15回 全國大會, p. 37, 1994.12.
4. B. T. Lowerre "HARPY speech recognition system," PhD thesis, Canegie-Mellon University, 1976.
5. John R. Deller, Jr., John G. Proakis and John H. L. Hansen, "Discrete-Time Processing of Speech Signals," Macmillan Publishing Company, 1993.
6. 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, 1995.5.
7. Shi-Wook Lee, Deuk-Sou Kim and Hyun-Yeol Chung, "A Korean Address Input System Employing Automatic Speech Recognition," Proc. ICSP '97, Vol. 1, No. 1, pp. 373-377, Aug. 1997.