

# 신뢰도 벡터 기반의 다단계 음성인식

전형배, 황규웅, 정 훈, 김승희, 박 준, 이운근(ETRI)

## <차 례>

- |                             |                        |
|-----------------------------|------------------------|
| 1. 서 론                      | 3. 신뢰도 벡터 기반의 다단계 음성인식 |
| 2. 다단계 음성인식 방법              | 3.1. 신뢰도 벡터            |
| 2.1. Acoustic Decoder       | 3.2. Cost 정의           |
| 2.2. Lexical Decoder        | 4. 실험결과                |
| 2.3. Lexical Model Training | 4.1. POI 인식실험          |
|                             | 5. 결 론                 |

## <Abstract>

### Multi-stage Speech Recognition Using Confidence Vector

Hyungbae Jeon, Kyuwoong Hwang, Hoon Chung, Seunghi Kim,  
Jun Park, Yunkeun Lee

In this paper, we propose a use of confidence vector as an intermediate input feature for multi-stage based speech recognition architecture to improve recognition accuracy. A multi-stage speech recognition structure is introduced as a method to reduce the computational complexity of the decoding procedure and then accomplish faster speech recognition. Conventional multi-stage speech recognition is usually composed of three stages, acoustic search, lexical search, and acoustic re-scoring. In this paper, we focus on improving the accuracy of the lexical decoding by introducing a confidence vector as an input feature instead of phoneme which was used typically. We take experimental results on 220K Korean Point-of-Interest (POI) domain and the experimental results show that the proposed method contributes on improving accuracy.

\* Keywords : Multi-stage speech recognition, Phoneme recognition, Confidence vector.

## 1. 서 론

음성인식 기술의 발달과 함께 최근에는 다양한 상용화 서비스가 시도되고 있다. 음성인식 기술의 상용화에 있어서 가장 큰 걸림돌이 되고 있는 것은 잡음 환경에서 음성인식 성능이 저하되는 것과, 최근 많이 요구되는 휴대용 단말기에서 대용량의 음성인식이 힘들다는 것이다. 최근 텔레매틱스 단말기 상에서의 point-of-interest (POI) 인식 분야가 음성인식의 상용화 분야로 주목 받고 있다. 그러나 POI 인식은 대상 어휘가 수십만 단어 이상이기 때문에 단말기에서 구동하는 것이 쉽지 않다. 그렇기 때문에 다양한 접근방법을 시도할 수 있는데, 그 중 하나가 다단계 인식기반 방법이다[1][2][4][5].

다단계 인식기반 음성인식은 음향학적 탐색과 언어적 탐색을 분리하는 방법이다. 음성신호의 특징벡터에서부터 음소인식을 수행하고, 인식된 음소 열을 사용하여 언어적 탐색 단계에서 단어 인식을 수행하는 것이다. 이와 같은 다단계 인식방법은 매우 가벼운 엔진인 음소인식기를 통해 음소인식결과를 특징벡터로 전달하게 되는데, 한 음소구간에서의 MFCC 특징 벡터 열이 하나의 음소 인덱스로 대치되어 특징벡터의 정보량을 극단적으로 줄일 수 있다. 인식된 음소 열을 언어적 탐색 단계인 lexical 인식단계의 특징으로 사용함으로써 계산량을 줄일 수 있는 것이다.

다단계 음성인식의 또 다른 장점은 휴대용 단말기에서는 매우 가벼운 음소인식기만을 수행시키고, 인식된 음소 열을 서버에 전송해 보다 복잡한 단어 인식을 수행할 수 있는 것이다. 이와 같은 음소인식 기반 분산처리 음성인식 방법이 최근에 소개되고 있다. 그러나 다단계 인식기는 음소인식기의 성능에 매우 영향을 많이 받게 되는 문제가 발생한다. 이를 극복하기 위해 더욱 정확한 음소인식기 개발이 요구되어 진다. 더욱 정확한 음소인식 방법으로 triphone 기반 음소인식 방법이 제안된 바 있다.

본 논문에서는 이와 같은 다단계 인식방법의 문제점을 극복하기 위해 음소인식기에서 lexical 인식기로 전달하는 특징벡터를 각 음소 구간에서의 신뢰도 벡터로 사용하는 것을 제안하였다. 음소인식기의 최적 음소 열 대신에 신뢰도 벡터 열을 lexical 인식기에 전달할 경우 lexical 인식기에서 바라보는 특징벡터의 정보가 더 늘어나게 되어, 더욱 정확한 인식을 할 수 있게 된다.

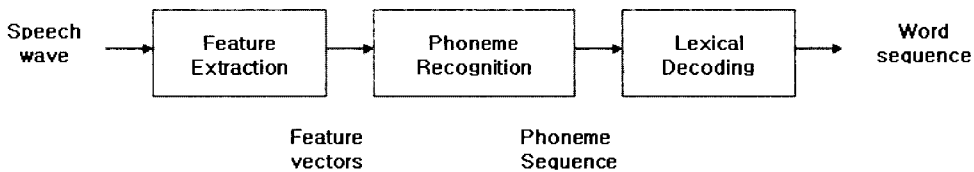
본 논문에서는 ETRI에서 개발하고 있는 다단계 인식기반 음성인식기의 기본구조를 제 2장에서 설명하고, 본 논문에서 제안한 방식을 제 3장에서 설명한다. POI 인식 실험 결과를 제 4장에서 설명하고, 제 5장에서 결론을 맺는다.

## 2. 다단계 음성인식 방법

### 2.1. Acoustic Decoder

다단계 음성인식은 크게 2단계로 구성된다. 음성신호가 입력될 경우 음소인식기에서 end-point detection (EPD)과 특징추출이 이루어지고, 음향학적 탐색 단계인 음소인식을 수행한다. 이후 언어적 탐색단계인 lexical 인식기에서 음소인식기에서 출력하는 음소 열을 기반으로 단어인식 또는 연결단어 인식을 수행한다.

다단계 음성인식의 구조도는 <그림 1>과 같다. 음소인식 단계의 음향모델은 일반적인 음성인식기의 음향모델을 사용한다. ETRI 음소인식기에서는 triphone 기반 음소인식기를 사용하였다[1]. 음소인식기는 휴대용 단말기에서도 동작할 수 있도록 메모리와 계산량 측면에서 최적화 작업을 수행하였다. 음소인식기의 메모리 최적화 작업은 탐색 단계에서 정적으로 고정되는 탐색 네트워크 관련 정보를 바이너리 이미지 형태로 미리 만드는 작업과 변수들을 관리하는데 있어 인덱스만을 사용하는 방법을 통해 이루어 졌다. 고속화를 위해서 해당 state의 Gaussian 중에서 입력된 특징벡터와 가장 가까운 하나의 Gaussian만을 사용하여 유사도 계산을 수행한다.



<그림 1> 다단계 음성인식 구성

### 2.2. Lexical Decoder

언어적 탐색 단계인 lexical 인식 단계에서는 음소인식 결과를 기반으로 단어 인식을 수행한다. 이때 음소의 치환 오류 확률을 모델링하는 lexical 모델을 사용하여 탐색을 수행한다. Lexical 탐색은 dynamic time warping (DTW) 방식으로 구현하였다.

DTW에 의한 탐색은 탐색 대상 단어를 y축에 음소 단위로 나열하고, 음소인식 결과를 x축에 음소 단위로 나열한 이후 DTW를 통해 구한 최적의 경로의 비용(cost)이 최소가 되는 단어를 인식 결과로 선택하는 것이다. 이 때 각 DTW 노드는 3개의 경로로 부터 천이되어 들어오는데, 각 경로를 치환(substitution), 삽입(insertion), 삭제(deletion) 음소 오류를 나타낸다. DTW 노드마다 3개의 천이 경로

중 비용(cost)이 가장 작은 경로를 선택한다. 각 경로의 비용은 각 경로의 오류 유형에 따라 미리 훈련된 음소 오류 모델로 부터 얻는다.

### 2.3. Lexical Model Training

Lexical 탐색 단계에서 음소 오류 모델을 기반으로 탐색을 수행하는데, lexical 탐색단계에 사용되는 확률모델이기 때문에 lexical 모델이라고 부르겠다. 음소 오류 모델 훈련은 주어진 음소 오류 모델 훈련 DB에 대해서 음소인식을 수행하고, 음소인식 결과를 바탕으로 훈련을 수행한다.

음소 오류 모델 훈련과정은 다음과 같다. 음소인식기의 음소인식 결과 음소 열과 정답 음소 열에 대해서 DTW 기반의 음소 열 정렬을 수행한다. 음소 열 정렬을 수행하면 주어진 참조 음소에 대해 음소인식기가 다른 음소로 오인식되는 치환 오류, 특정 참조 음소가 음소인식 결과에서 나타나지 않는 삭제 오류, 음소인식 결과 중 특정 음소가 추가로 관측되는 삽입 오류의 3가지 오류를 관측할 수 있다. 이와 같은 3가지 오류를 모든 훈련 DB에 대해서 누적하고, 누적된 경우의 수로 부터 각 오류 유형의 확률을 계산한다. 음소 오류 확률 모델이 업데이트되면 DTW를 통한 음소 열 정렬 과정에 음소 오류 확률 모델을 적용하여 더욱 정확한 음소 열 정렬을 수행하게 된다. 이와 같이 음소 열 정렬과 음소 오류 모델 업데이트를 반복적으로 수행함으로써 음소 오류 모델 훈련을 수행한다.

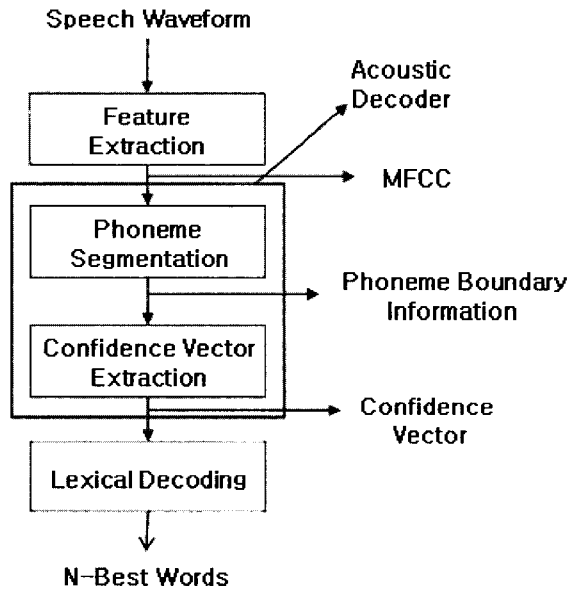
## 3. 신뢰도 벡터 기반 다단계 음성인식

### 3.1. 신뢰도 벡터

본 논문에서 제안하는 다단계 음성인식은 다음의 3단계로 구성되어 있다. 제 1 단계에서는 음소인식을 통해 최적의 음소 열의 음소 구간을 찾는다. 이는 음소인식 결과 중 최적의 음소 열 결과의 음소 경계 정보를 사용하였다. 제 2단계에서는 제 1단계에서 구한 음소 경계 내에서 각 음소의 확률을 계산한다. 음소 확률은 일반적인 비터비(Viterbi) 탐색을 통해 얻은 유사도 값으로 부터 정규화과정을 통해 정의한다. 식 (1)에서 본 논문에서 정의한 신뢰도를 설명하였다. 비터비 탐색을 통해 유사도를 계산하고,  $i$ 번째 음소의 유사도 값과, 정의된 전체  $N$ 개 음소의 유사도 값의 합과의 비율로써 현재의 음소구간이  $i$ 번째 음소일 확률을 정의한다. 이와 같이 구한 확률 값에 로그를 취해서 각 음소의 신뢰도 값으로 사용한다. 식 (1)에서  $i, j$ 는 각 음소의 인덱스를 나타내고  $N$ 은 정의된 음소의 개수를 의미한다. 본 논문에서는 한국어 음소를 묶음이 포함된 46개 음소로 정의하여 사용하였다. 각 단계

를 간단히 보이면 <그림 2>와 같다.

$$confidence(i) = \log \left( \frac{likelihood(i)}{\sum_{j=1}^N likelihood(j)} \right). \tag{1}$$

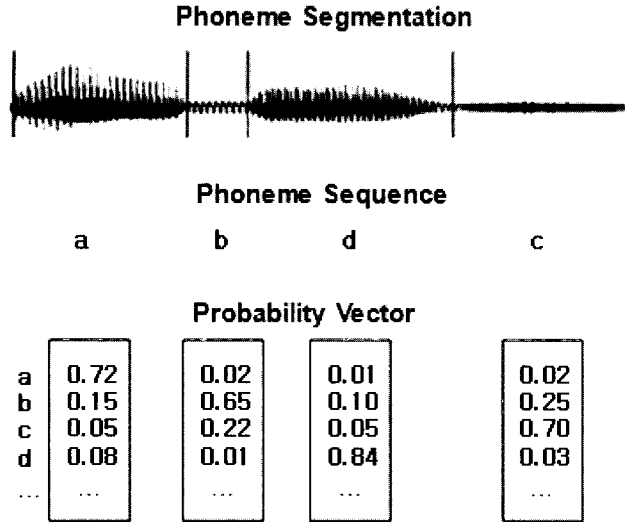


<그림 2> 신뢰도 벡터 기반 다단계 음성인식 구성

음소구간과 신뢰도 벡터추출 과정의 음소인식기는 앞에서와 동일하게 triphone 기반 음소인식기를 사용하였다. triphone 기반 음소인식기에서의 신뢰도의 정의는 center-phone이 동일한 모든 triphone 중에서 가장 유사도 확률이 좋은 음소를 해당 center-phone의 대표 음소로 선택하여 식 (1)에 적용하여 계산하였다.

신뢰도 벡터는 lexical 인식 단계에서 특징벡터로서 사용되어 진다. 기존의 다단계 음성인식 방법에서는 <그림 3>에서의 음소 열 “a b d c”가 특징 데이터로 사용되어 lexical 탐색을 수행한다. 음소인식 결과 음소 열과 미리 훈련된 음소 오류 모델을 기반으로 참조 단어의 음소 열과의 정렬 비용을 계산하여 탐색을 수행한다.

본 논문에서 제안하는 신뢰도 벡터 기반 다단계 음성인식은 음소인식 결과의 음소 열 대신에 특징 벡터 열을 lexical 탐색단계의 특징 데이터로 사용한다. 음성 파형에 대한 음소구간과 음소인식 결과 음소 열, 해당 음소 구간에서의 각 음소의 확률 벡터의 예를 <그림 3>에 나타내었다.



<그림 3> 음소 구간과 신뢰도 벡터

Lexical 탐색을 위하여 음소 열 기반 다단계 음성인식 방법의 음소 오류 모델과 유사한 lexical 모델이 필요하다. lexical 모델은 임의의 훈련 DB로 부터 구한 신뢰도 벡터들과 정답 음소 열을 기반으로 훈련을 수행한다. 기본적인 훈련 과정은 2.3절에서 소개한 lexical 모델 훈련 과정과 유사하다. 신뢰도 기반 lexical 모델은 DTW 정렬에서의 비용의 정의를 식 (2)와 같이 새로 정의하여 사용한다. 식 (2)에서 정의한 비용 정의를 사용하여 음소인식 결과 신뢰도 벡터 열과 참조 음소 열을 정렬하여, 참조 음소에 대해서 신뢰도 벡터 값을 누적한다. 훈련 DB 전체에 대해 각 음소들의 신뢰도 벡터 값 누적을 마치면, 각 음소의 평균 신뢰도 벡터 값을 계산하여 업데이트하고, 이 값을 식 (2)의 가중치 벡터로 사용하는 것이다.

음소 열 기반 다단계 음소인식 방법에서와 다르게 본 논문에서 제안하는 신뢰도 기반 다단계 음성인식 방법은 음소구간(segment)에서 음소들의 신뢰도를 언어적 탐색 단계의 특징벡터로 사용한다. 신뢰도 기반 다단계 음소인식 방법은 음소구간에 대해 1개의 음소 인덱스를 전달하는 대신에 음소들의 신뢰도 벡터를 전달하기 때문에 lexical 탐색단계에서 사용하는 정보가 늘어나게 되고, 음성인식 성능이 환경적 요인에 더욱 강인해 질 수 있게 된다.

### 3.2. Cost 정의

Lexical 탐색 단계에서는 참조 음소에 대한 평균적 신뢰도 분포가 입력 특징벡터의 가중치가 되어, 입력 신뢰도 벡터에 대해 가중치 합을 수행하면 해당 음소에 대한 관측 신뢰도 벡터의 확률 값이 된다. 확률 값은 변동폭을 고려하여 로그로

변환하여 비용(cost)으로 사용하며 비용의 정의는 식 (2)와 같다.

$$\text{cost}(f|p) = -\log\left(\sum_{i=1}^N (\exp(w_p(i)) \times \exp(f(i)))\right). \quad (2)$$

식 (2)는 신뢰도 벡터  $f$ 가 음소  $p$ 일 비용을 정의한다. 음소  $p$ 에 대한 평균적 신뢰도 분포를 가중치 벡터  $w_p$ 로 정의하고, 이를 모든 원소에 대해 신뢰도 벡터와 가중치 합 계산을 수행한다. 식 (2)의 비용(cost)은 DTW 방식의 lexical 탐색에서의 치환 오류 경로의 비용(cost)으로써 사용되어 진다.

## 4. 실험 결과

### 4.1. POI 인식 실험

제한한 방법의 성능을 검증하기 위해 고립단어 인식실험과 finite state network (FSN) 기반 연결단어 인식실험을 수행하였다. POI 인식 실험은 22만 어휘 POI를 인식 대상으로 한다. 평가데이터는 7명의 화자로 부터 수집한 총 693 발화이다. 평가데이터는 POI 22만 어휘 중 임의의 100개의 어휘를 발성한 것이다. 각 음성 신호는 위너(Wiener) 잡음처리를 수행하였다.

음소인식기는 MFCC를 특징으로 하고, triphone을 음향모델로 사용하였다. 음향 모델 훈련은 실주행 차량환경에서 수집한 텔레메틱스 음성 DB를 사용하였다. 텔레메틱스 음성 DB는 ETRI에서 배포하는 2004년도, 2005년도 자동차 환경에서의 음성명령어 및 정보 검색용 대화체 문장 DB이다. Triphone 기반 음소인식기는 점유 메모리가 약 7 MByte이고, 3 GHz dual-core CPU에서 실시간 성능 0.03 정도의 인식 속도를 갖는다.

Lexical 인식기에서의 lexical 모델을 위한 훈련 셋이 필요하게 된다. 주어진 평가 DB가 적은 관계로 기본적인 실험은 closed 평가실험을 하였다. Closed 실험은 인식하고자 하는 7명의 발화로 부터 lexical 모델을 훈련하는 것이다.

Open 실험을 위하여 N-fold 평가 실험을 수행하였다. 즉, 7명의 화자 음성파일 중, 6명의 음성파일로 부터 lexical 모델을 훈련하고, 나머지 1명의 음성파일로 평가실험을 수행하였다. 이와 같은 방식으로 총 7번 실험을 수행하였다.

기존의 음소 열을 특징으로 사용하는 다단계 음성인식 방법과 제한한 신뢰도 벡터를 특징벡터로 사용하는 다단계 음성인식 방법의 성능 차이를 자세히 관찰하기 위하여 N-Best 인식결과에 대해서 비교 분석을 수행하였다.

<표 1>은 다단계 음성인식 방법의 POI 평가 데이터 실험 결과인 단어 인식을 이다. 음소인식기 음소 천이 규칙은 한국어의 음절 규칙을 지키면서 triphone 음소

의 문맥정보를 유지하는 것이다. 이와 같은 음소 천이 제한 조건이 존재하더라도 실제로 음소인식 실험을 수행할 경우 음소들의 삽입 오류가 많이 발생하게 된다. 이와 같은 삽입 오류를 조절하기 위하여 phone insertion penalty (PIP)를 정의하여 음소 천이 시에 penalty 값으로 적용하였다. 정의한 PIP 값에 따라 음소 인식을 차이가 발생하여 <표 1>에서와 같이 여러 PIP 값에 대해 실험을 수행하였다. 위 실험에서 사용한 lexical 모델은 평가 데이터로 부터 훈련한 closed 평가 결과이다.

<표 1> Phone Insertion Penalty (PIP)에 따른 POI 인식 실험 (Closed Test)

| PIP | 단어 인식률 (%)   |              |
|-----|--------------|--------------|
|     | 음소 열 방식      | 신뢰도 벡터 방식    |
| 4   | 75.76        | 77.77        |
| 10  | 75.76        | <b>77.92</b> |
| 15  | 75.61        | 77.77        |
| 20  | 76.33        | 77.21        |
| 25  | <b>77.06</b> | 77.05        |
| 30  | 75.61        | 75.18        |
| 35  | 72.87        | 73.73        |

<표 1>에서 보는 것과 같이 제안한 신뢰도 벡터를 특징벡터로 사용하는 다단계 음성인식 방식이 최고 성능으로 0.9% 가량 더 좋았으며, PIP의 변화에 비교적 무관하게 좋은 성능을 내는 것을 확인할 수 있었다.

<표 2>는 N-Best 성능평가를 결과를 나타낸다. 각 실험은 <표 1>의 실험에서 최고의 성능을 낸 PIP 값을 사용하여 실험을 수행한 것이다. 실험결과를 보면 제안한 신뢰도 벡터 방식의 다단계 음성인식 방법이 N-Best 결과에서 성능차이는 더욱 더 벌어지고 있는 것을 확인할 수 있었다. 2~5 Best 결과에서 3~4% 가량의 성능향상을 얻을 수 있었다.

<표 3>은 open 평가 실험을 하기 위하여 N-fold 방식의 실험을 한 경우의 성능을 정리한 것이다. N-fold 방식의 평가이기 때문에 <표 2>의 closed 평가 실험에 비해 성능이 저하되는 것을 확인할 수 있었다. 그러나 제안한 방식의 성능저하는 음소 열을 특징벡터로 사용하는 다단계 인식기에 비해 성능저하 정도가 적은 것을 알 수 있다. 두 방식의 성능저하 정도를 <그림 4>에 나타내었다. <그림 4>에서 x축은 N-best를 의미하고, y축은 closed 평가에 대해 N-fold 평가에 의한 성능저하 %를 나타낸다. 각 N-best 성능을 살펴보면 음소 열 방식에 비해 성능저하 폭이



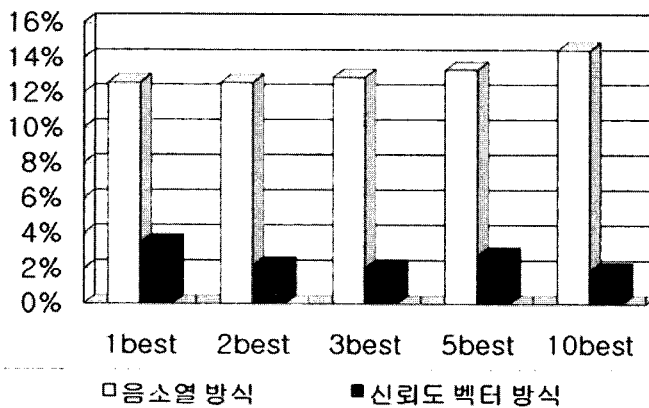
1/4 이하인 것을 확인할 수 있었다.

<표 2> N-Best 성능평가 (Closed Test)

| N-best  | 단어 인식률 (%) |           |
|---------|------------|-----------|
|         | 음소 열 방식    | 신뢰도 벡터 방식 |
| 1 Best  | 77.06      | 77.92     |
| 2 Best  | 81.67      | 85.40     |
| 3 Best  | 85.43      | 88.44     |
| 5 Best  | 88.89      | 91.33     |
| 10 Best | 93.07      | 94.36     |

<표 3> N-Best 성능평가 (N-fold Test)

| N-best  | 단어 인식률 (%) |           |
|---------|------------|-----------|
|         | 음소 열 방식    | 신뢰도 벡터 방식 |
| 1 Best  | 64.50      | 74.46     |
| 2 Best  | 69.12      | 83.12     |
| 3 Best  | 72.58      | 86.29     |
| 5 Best  | 75.61      | 88.46     |
| 10 Best | 78.64      | 92.35     |



<그림 4> Closed 성능평가 대비 N-fold 성능평가의 성능 저하 정도 비교

<표 2>와 <표 3>에서의 결과로 부터 lexical 인식단계에서 음소 열만을 특징벡터로 사용하는 것에 비해, 각 음소구간에서의 음소들의 신뢰도 벡터를 전달받는 경우의 다단계 음성인식기가 더욱 우수한 것을 확인하였다. 이 결과는 N-fold의 open 평가 실험에서도 일관되게 나타나고 있다. 즉 음소인식기에서 lexical 인식기로 더욱 많은 정보를 전달한 경우에 전체적으로 인식을 향상을 얻을 수 있었으며, 또한 음성인식기 성능도 더욱 강인해지는 것을 확인할 수 있었다.

<표 4> 점유 메모리량과 속도 비교

| 메모리/속도           | 음소 열 방식 | 신뢰도 벡터 방식 |
|------------------|---------|-----------|
| 메모리량 (MByte)     | 80      | 85        |
| Real-time Factor | 0.29    | 0.27      |

<표 4>에서는 음소 열 기반의 다단계 음성인식 방법과 신뢰도 벡터 기반의 다단계 음성인식 방법의 점유 메모리량과 인식 속도를 비교한 것이다. 신뢰도 벡터 기반의 다단계 음성인식 방법은 음소인식기를 음소 구간 검출과 신뢰도 벡터 추출에 사용하기 때문에 메모리 요구량이 약간 증가하게 된다. 점유 메모리량은 동적, 정적 메모리를 합한 것으로 정적 메모리는 약간 증가하지만, 동적 메모리가 감소하였다. 속도 측면에서 신뢰도 벡터 추출 과정이 추가되어 계산량이 늘어나지만 lexical 탐색 단계에서 활성 노드의 개수가 약 1/10로 줄어들어 전체적인 인식 속도가 향상되는 것을 확인하였다. 이와 같이 활성 노드의 개수가 줄어들면서 인식 성능이 향상되는 것으로 부터 lexical 탐색단계에서의 특징 벡터로서 신뢰도 벡터가 음소 열보다 더욱 변별력이 강한 특징인 것을 알 수 있었다.

## 5. 결론

본 논문에서는 새로이 제안되고 있는 다단계 음성인식에 대하여 실험하였다. 다단계 베이스라인 음성인식기에서는 음소인식기에서 최적의 음소 열을 lexical 인식기에 전달한다. 반면 본 논문에서는 음소인식기에서 음소 구간에서의 각 음소의 신뢰도 벡터를 lexical 인식기에 전달한다. 이와 같이 음소인식기와 lexical 인식기 사이의 정보량을 확대함으로써 N-best 성능과 N-fold 성능에서 성능향상을 얻을 수 있었고 보다 강인한 성능을 나타내었다.

제안한 다단계 음성인식은 lexical 모델이 신뢰도 벡터의 분포로서 모델링되기 때문에 일반적인 MAP 기반 적응방법을 도입할 수 있을 것이다. 이와 같은 적응을 통해 실제 서비스될 수 있는 음성인식 엔진으로의 가능성을 높여줄 것이다.

## 참 고 문 헌

- [1] 김승희, 황규웅, 전형배, 정훈, 박준, “분산음성인식을 위한 내장형 고속/경량 음소인식기 개발”, *한국정보처리학회 춘계학술발표대회논문집*, 제14권, pp. 395-396, 2007.
- [2] V. Zue, J. Glass, D. Goodine, M. Phillips, S. Seneff, “The SUMMIT speech recognition system: Phonological modelling and lexical access”, *Proc. ICASSP*, pp. 49-52, 1990.
- [3] 정익주, 정훈, “임베디드용 대용량 음성인식 시스템 구현”, *2006년도 한국음향학회 학술대회 논문집*, 제25권, 제2(s)호, pp. 151-154, 2006.
- [4] K. Demuynck, D. V. Comperolle, H. V. Hamme, “Robust phone lattice decoding”, *Proc. ICSLP*, pp. 1622-1625, 2006.
- [5] L. Ten Bosch, A. Hamalainen, O. Scharenborg, L. Boves, “Acoustic scores and symbolic mismatch penalties in phone lattices”, *Proc. ICASSP*, pp. 437-440, 2006.

접수일자: 2007년 5월 17일

게재결정: 2007년 9월 3일

▶ 전형배(Hyungbae Jeon) : 교신저자

주소: 305-350 대전광역시 유성구 가정동 161번지

소속: 한국전자통신연구원(ETRI) 음성/언어정보연구센터 음성처리연구팀

전화: 042) 860-5788

E-mail: hbjeon@etri.re.kr

▶ 황규웅(Kyuwoong Hwang)

주소: 305-350 대전광역시 유성구 가정동 161번지

소속: 한국전자통신연구원(ETRI) 음성/언어정보연구센터 자동통역연구팀

전화: 042) 860-3941

E-mail: kyuwoong@etri.re.kr

▶ 정 훈(Hoon Chung)

주소: 305-350 대전광역시 유성구 가정동 161번지

소속: 한국전자통신연구원(ETRI) 음성/언어정보연구센터 음성처리연구팀

전화: 042) 860-5742

E-mail: hchung@etri.re.kr

▶ 김승희(Seunghi Kim)

주소: 305-350 대전광역시 유성구 가정동 161번지

소속: 한국전자통신연구원(ETRI) 음성/언어정보연구센터 자동통역연구팀

전화: 042) 860-6187

E-mail: seunghi@etri.re.kr

## ▶ 박 준(Jun Park)

주소: 305-350 대전광역시 유성구 가정동 161번지

소속: 한국전자통신연구원(ETRI) 음성/언어정보연구센터 자동통역연구팀

전화: 042) 860-6934

E-mail: junpark@etri.re.kr

## ▶ 이윤근(Yunkeun Lee)

주소: 305-350 대전광역시 유성구 가정동 161번지

소속: 한국전자통신연구원(ETRI) 음성/언어정보연구센터 음성처리연구팀

전화: 042) 860-1869

E-mail: ykleee@etri.re.kr