Journal of Korea Multimedia Society Vol. 13, No. 4, April 2010(pp. 496-503)

음향학적 및 언어적 탐색을 이용한 어휘 인식 최적화

안찬식[†], 오상엽[††]

요 약

어휘인식 시스템은 스탠드 얼론(Standalone)으로 개발되어 지고 있으며 휴대용 단말기에서 사용하였을 경우 메모리 공간의 제약과 오디오 압축으로 인해 인식률이 낮게 나타난다. 본 연구에서는 휴대용 단말기의 성능과 인식률 향상을 위하여 음향학적 탐색과 언어적 탐색을 분리하여 어휘 인식 속도를 개선한 시스템을 제안하였다. 음향학적 탐색은 휴대용 단말기에서 수행하고 보다 복잡한 언어적 탐색은 서버에서 처리하는 시스템으로 음성신호로부터 특징벡터를 추출하여 GMM을 이용한 음소인식을 수행하고, 인식된 음소 열을 서버로 전송하여 렉시컬 트리 탐색 알고리즘을 사용하여 언어적 탐색 단계에서 어휘 인식을 수행하였다. 시스템 성능 평가 결과 어휘 종속 인식률은 98.01%, 어휘 독립 인식률은 97.71%의 인식률을 나타냈으며 인식속도는 1.58초로 나타내었다.

The Vocabulary Recognition Optimize using Acoustic and Lexical Search

Chan Shik Ahn[†], Sang Yeob Oh[††]

ABSTRACT

Speech recognition system is developed of standalone, In case of a mobile terminal using that low recognition rate represent because of limitation of memory size and audio compression. This study suggest vocabulary recognition highest performance improvement system for separate acoustic search and lexical search. Acoustic search is carry out in mobile terminal, lexical search is carry out in server processing system. feature vector of speech signal extract using GMM a phoneme execution, recognition a phoneme list transmission server using Lexical Tree Search algorithm lexical search recognition execution. System performance as a result of represent vocabulary dependence recognition rate of 98.01%, vocabulary independence recognition rate of 97.71%, represent recognition speed of 1.58 second.

Key words: speech recognition(어휘인식), acoustic search(음향학적 탐색), lexical search(언어적 탐색)

1. 서 론

인터넷의 보편화와 어휘인식 기술의 발달로 컴퓨터 사용 환경에 많은 변화가 일고 있다. 개인용 PC에서 주로 이용하던 서비스가 훨씬 다양한 디바이스에서 상용화 서비스를 시도하고 있다. 어휘인식 기술의 상용화에 있어서 가장 큰 걸림돌이 되고 있는 것은 휴대용 단말기에서의 대용량의 어휘인식과 잡음 환경에서 어휘인식 성능이 저하된다. 현재까지의 어휘인식 모델은 스탠드 얼론으로 개발되어 휴대용 단말기의 프로세싱으로 처리하기에는 메모리 공간의 제약과 오디오 압축으로 인한 인식률 저하와 한정적인

※ 교신저자(Corresponding Author) : 오상엽, 주소 : 경기도 성남시 수정구 복정동(461-702), 전화 : 031)750-5798, FAX: 02)426-9159, E-mail : syoh@kyungwon.ac.kr
접수일 : 2009년 9월 19일, 수정일 : 2009년 11월 16일
완료일 : 2010년 1월 21일

[†] 정회원, 광운대학교 컴퓨터공학과
(E-mail : csan1004@paran.com)
[††] 종신회원, 경원대학교 IT대학 컴퓨터소프트웨어 교수
※ 본 연구는 2010년도 경원대학교 지원에 의한 결과임

용량의 인식 수행으로 나타난다. 또한 무선 통신환경이 발달하고 있지만 음성신호를 직접 전송하기에는 대역폭에 대한 한계가 존재한다[1,2]. 이를 극복하기 위해 휴대용 단말기와 서버를 연계하여 처리하는 분산 어휘인식 시스템의 개발이 요구되어지며 현재 다양한 방법의 분산 어휘인식 시스템이 개발되어 지고 있다[3,4].

분산 어휘인식 시스템은 휴대용 단말기에서는 간단한 음소인식만을 수행하고 인식 결과를 서버로 전송하여 복잡한 단어 인식을 수행하는 방법이다. 기존의 시스템에서는 통계적 방법에 의한 어휘인식을 수행하여 N-gram을 이용한 통계적 문법으로 인식하는 시스템을 구축하였다. 그러나 통계적 방법을 사용하는 것은 이미 가지고 있는 확률에 의존해야 하므로 언어모델을 구하기 위해 많은 양의 말뭉치가 필요하게 되고 메모리에 저장해야 될 정보가 기하급수적으로 늘어나게 되어 제한된 공간에서 사용할 수 없는 단점이 있다[5].

본 논문에서는 분산 어휘인식 시스템을 개발하기 위하여 음향학적 탐색과 언어적 탐색을 분리하여 각각의 시스템에서 처리하는 방법을 실험하였다. 어휘인식을 위한 음향학적 탐색은 휴대용 단말기에서 수행하고 보다 복잡한 언어적 탐색은 서버에서 처리하는 것으로 입력된 음성신호로부터 특징벡터를 추출하여 GMM(Gaussian Mixture Model)을 이용한 음소인식을 수행하고, 인식된 음소 열을 서버로 전송하여 렉시컬 트리 탐색을 수정한 알고리즘을 사용하여 언어적 탐색 단계에서 단어 인식을 수행한다. 시스템 성능 평가 결과 어휘 종속 인식률은 98.01%, 어휘 독립 인식률은 97.71%의 인식률을 나타냈으며 인식 속도는 1.58초의 결과를 얻을 수 있었다. 이는 스탠드얼론에서의 결과 값과 비교하였을 때 속도 면에서 0.3초 향상되었으며 인식률에서 1.1% 향상된 결과였다. 또한 인식 단어가 늘어났을 때는 오히려 더 좋은 성능을 나타냈다.

제 2장에서는 일체형 구조의 인식 시스템과 분산 구조의 인식 시스템에 대하여 간략히 소개하고, 제 3장에서는 본 논문에서 제안한 분산 어휘 인식 시스템에 대하여 설명한다. 제 4장에서는 제안한 시스템의 실험결과에 대하여 설명하고 제 5장에서 결론을 맺는다.

2. 관련연구

2.1 일체형 어휘 인식

일체형 어휘 인식 구조는 탐색 과정에서 모든 가능한 지식 정보들을 가져오며 복잡한 언어 모델의 기초위에 제작된 단어 그래프를 통하여 음향 모델과 언어 모델을 사용한다. 이와 같은 방식의 장점은 검색의 효율성에 있다. 음성에서 음향적 혼란이 많아 어휘 및 언어 모델에서 제공되는 정보를 일찍 포함하는 것은 탐색 공간으로부터 가장 가능성이 낮은 부분들을 삭제하기 위해 효과적이다. 그러나 이러한 일체형 검색 방법은 몇 가지 단점을 가지고 있다. 첫째, 음향 모델의 left-to-right 동작은 왼쪽에서 오른쪽으로 동작하기 위해 다른 정보들을 요구하게 되는데, 종종 오른쪽 문맥을 의존하는 언어 모델의 경우, 부분적으로 비효율적인 의사 결정을 야기한다. 둘째, 일반 음성인식 구조에서의 새로운 정보의 삽입이 어렵다. 그림 1은 일반적인 음성인식 시스템의 블록 다이아그램을 보여준다. 이것은 특징 추출과 패턴 인식기라는 2개의 콤포넌트로 구성된다. 특징 추출기는 입력되는 음성에서 인식을 위한 패턴 인식기에서 사용되어 지는 입력음성으로 부터 MFCC(Mel-frequency cepstral coefficients), PLP(Perceptual Linear Prediction), LPCC(linear predictive cepstrum coefficients)와 같은 특징을 계산한다[6-8].

2.2 분산 어휘 인식

초창기의 분산 어휘 인식 시스템은 클라이언트 단에서 음성만 받아들여 서버로 보내 인식을 하는데 음성의 데이터의 사이즈가 커서, 무선 통신 환경에서는 실시간 인식이 가능하지 못하여 유선 네트워크에서 가능한 인식 시스템이다.

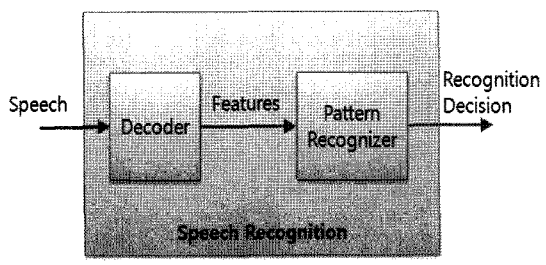


그림 1. 일체형 어휘인식 시스템

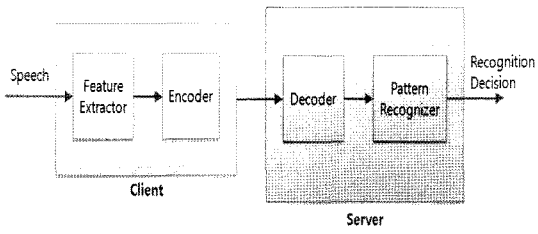


그림 2. 분산 어휘인식 시스템

현재 무선 통신환경이 계속적으로 발달하였지만 아직은 원음을 보내기에는 대역폭에 대한 한계가 존재한다. 이러한 문제점으로 인하여 클라이언트 단에서 음성을 입력받아 전처리하여 서버로 보낸다. 서버에는 클라이언트에서 특징추출을 끝낸 데이터를 받아 어휘 인식 단에서 인식하여, 그 결과를 다시 클라이언트로 보낸다.

이러한 경우 클라이언트 단에서의 특징을 추출하기위한 시간이 걸리지만, 실시간으로 처리하는 데는 문제가 없다.

그림 2는 클라이언트-서버의 분산 어휘 인식시스템이다. 여기서 클라이언트는 특징 추출기를 포함하여 추출된 특징은 인코딩하고, 디코딩된 특징 데이터를 가지고 어휘 인식을 수행하는 서버로 전송한다.

분산 시스템의 인코더는 주어진 비트률에서 최적의 인식 성능을 보장해야 하며 인식에서 GMM, G.721.1, G.727, G.728, G.729, ADPCM, MELP 등 다양한 음성 코딩 기술의 효과는 수많은 실험에 의해 평가되어진다[9].

3. 어휘인식 속도 향상을 위한 시스템 모델링

3.1 시스템 지원 DB

어휘 인식시스템을 구현하기 위하여 한국어의 음운현상의 최대화를 위하여 3종류의 단어음성 DB를 가지고 문맥 종속적 음향모델을 구축하였고, 최적의 CHMM(Continuous Hidden Markov Model)의 구조, 즉 상태수와 Gaussian Mixture의 개수를 결정하였으며, 3가지의 상태를 가지는 Left-to-Right의 상태를 갖는 유사음소단위의 문맥종속적인 Triphone을 바탕으로 한 인식시스템으로써 특징파라미터로 12차의 MFCC와 0차 MFCC, 1차의 에너지를 사용하여 총 14차의 특징을 추출하여 인식실험에 사용하였다. 가변어휘 인식 실험용 음소모델 훈련을 위하여 총 3종류(ETRI의 PBW445 DB, POW3848 DB, 국어공학연구소의 PBW452 DB)의 음성 DB를 사용하였고, 음성 DB는 8k로 Sampling하고 16bit 양자화한 선형 PCM의 포맷을 갖는다.

3.2 GMM을 이용한 음소인식

GMM은 출력 확률밀도함수가 가우시안 밀도혼합인 1개의 상태만으로 구성된 CHMM의 한 형태이다. 이러한 GMM은 다음과 같은 특징을 가지고 있다.

첫째, GMM은 음향학적 클래스의 집합을 모델링할 수 있다. 발성에 대응되는 음향공간은 모음이나 비음, 파찰음과 같은 음소를 표현하는 음향학적 클래스의 집합으로 잘 표현된다.

둘째, 단봉 가우시안 음소모델은 평균 벡터의 특징벡터와 공분산으로 각 음소의 특징벡터의 이산집합으로 음소분포를 표현한다. 이와 같은 점을 고려하여 구성된 GMM은 가우시안 함수의 이산집합을 사용하여, 각각의 평균과 공분산을 가지게 함으로써 이들 두 모델의 특징을 혼합한 형태이다[10].

가우시안 혼합 밀도는 M 성분 밀도의 가중합계로서 식 (1)에 의해 얻어진다.

$$p(x|\lambda)=\sum_{i=1}^{M} c_i b_i(x) \tag{1}$$

x는 d-차원 랜덤 벡터이며, $b_i(x)$, $i=1,\cdots,M$는 i번째의 성분 밀도이고, c_i, $i=1,2,\cdots,M$는 i번째 혼합 밀도 가중치이다. 각 혼합 밀도의 가중치는 다음과 같이 제한된다.

$$\sum_{i=1}^{M} c_i=1 \tag{2}$$

각 성분 밀도는 평균 μ_i과 공분산 Σ_i를 가지는 d-변량 가우시안 함수이다. 가우시안 혼합 밀도는 모든 성분 밀도의 혼합밀도 가중치와 공분산 행렬, 평균벡터로 구성된다. 따라서 GMM의 파라미터를 구하면 아래와 같은 모델을 만든다.

$$\lambda = c_i, \mu_i, \Sigma_i, \quad i=1,\cdots,M \tag{3}$$

모델 학습은 주어진 학습음성으로부터 학습특징벡터의 분포와 가장 잘 맞는 GMM 파라미터를 추정하는 것이다. GMM의 파라미터를 추정하는 방법에는 여러 가지가 있으나, 가장 잘 알려진 방법으로는 MLE(Maximum Likelihood Estimation)이다. MLE

는 주어진 학습데이터에서 GMM의 유사도를 최대화하는 모델파라미터를 찾는데 사용된다. T학습벡터 $X=x_1, x_2, \cdots, x_T$의 열에서, GMM 유사도는 다음과 같고,

$$P(X|\lambda) = \prod_{t=1}^{T} p(x_t|\lambda) \quad (4)$$

이를 로그영역에서 표현하면 다음과 같다.

$$L(X|\lambda) = \sum_{t=1}^{T} \log p(x_t|\lambda)$$

본 논문에서 제안하는 음소인식기에서는 입력음성과 모델과의 유사도를 GMM 확률값을 이용하여 계산하였다. GMM은 특정 파라미터의 기댓값이 가우시안 분포를 가진다고 가정하고 그에 의한 확률값을 도출한다. GMM은 평균과 표준편차만으로 값들에 대한 특징을 표현할 수 있기 때문에 널리 이용된다. GMM에 사용된 식은 식(5)와 같다.

이 때 d는 특징 파라미터의 차수를 나타내고, μ가우시안 모델의 평균을, Σ는 가우시안 모델의 공분산 매트릭스를 나타낸다[11,12].

GMM 음소 학습단계에서는 CHMM으로 구성된 음소 모델에 의한 자동 음소분할로 구한 라벨 정보를 이용하여 43개의 음소별 데이터베이스를 구축하고 이를 이용한 43개의 음소별 GMM 파라미터를 추정한다. 이후 음소 인식과정에서 음소별 GMM의 평균, 공분산과 CHMM의 중간상태 천이 확률을 이용한 연속 음소 인식 네트워크를 구성하고 이를 통해 최대 사후확률을 갖는 음소열을 발생한다. 발생된 음소열은 인코딩하여 서버로 전송된다.

3.3 렉시컬 트리 탐색 알고리즘을 이용한 언어적 탐색

인코딩하여 전송되어진 음소열을 디코딩하여 미리 구성된 탐색 네트웍에서 가장 최적인 경로를 찾는다. 탐색에 사용되는 네트웍은 음향모델, 발음사전, 언어모델을 결합하여 하나의 커다란 네트웍으로 구성되며, 탐색 네트웍은 소규모 어휘 인식에서는 유한상태망을 사용하나 본 논문에서는 대어휘 인식을 위한 통계적 언어모델이 적용된 렉시컬 트리 탐색을 사용하였다. 언어모델의 기본 단위는 의사형태소를 가정하였고 언어모델은 트라이그램을 사용하였다. 발음사전은 일반적으로 한 단어가 발음되는 발음열을 나열하지만 여기에서는 한 단어가 여러 개의 발음을 가질 수 있으며 이를 그래프로 표현한다고 가정하고 인식된 결과는 1-best 인식결과와 lattice 형태의 인식결과를 얻을 수 있으며 lattice 형태의 인식결과로부터 N-best의 인식결과를 얻을 수 있다.

렉시컬 트리 탐색의 일반적인 알고리즘은 다음과 같다.

```
void BinarySrearch(int A[], int First, int Last, int
Key, int& Index)
{
    if (First >Last)
        index = -1;
    else
    {
        int Mid = (First+Last)/2;
        if(Key ==A[Mid])
            Index = Mid;
            else if(Key < A[Mid])
                BinarySrearch(A,First, Mid-1,Key,
                Index);
            else

BinarySrearch(A,Mid+1,Last,Key,Index);
    }
}
```

변형한 렉시컬 트리 탐색은 그림 3과 같은 형태의 인식네트워크를 가지는 구조이다. 인식해야하는 어휘의 수가 많을 경우와 연속음성인식을 위한 탐색구조는 일반적으로 발음사전을 트리의 형태로 구성한 렉시컬 트리 탐색을 사용한다. 어휘는 x, y, z의 세 단어로 구성되고 발음은 $x=ab$ 또는 $x=af, y=acd, z=ace$로 가정하였을 때 주어진 렉시컬 트리 탐색은 x, y, z의 조합으로 이루어진 임의의 단어 열을 인식한다.

렉시컬 트리 탐색의 장점은 음향 문맥이 동일한 노드들을 공유함으로써 탐색공간을 줄일 수 있다[13].

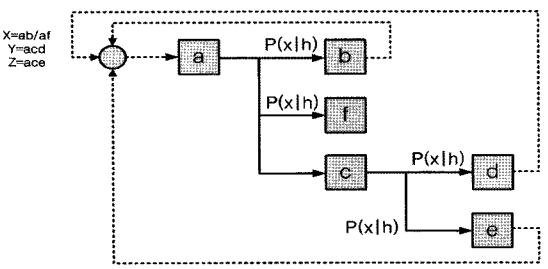


그림 3. 렉시컬 트리 탐색

렉시컬 트리 탐색을 변형한 알고리즘은 다음과 같다.

```
function LexicalTreeSearch
for all states
   Accumulate likelihood.
     like(j,t)=like(i,t-1)+log aij+log b(j,xt)
end

for all penultimate phone nodes
        Add LM score.
end

for all leaf nodes
   Consider transitions from leaf nodes to root node.
        like(j,t)=like(end,t)+Bigram(wi,wt)
   Record the previous node, entering time, like-
lihood for backtracking.

end
end
```

변형된 알고리즘은 모든 프레임에 대하여 모든 상태노드에서 단어내 천이 규칙을 적용하고 모든 잎사귀노드 하나 이전의 노드에 대하여 언어 모델값을 적용한다. 잎사귀 노드에서 루트노드로 가는 가장 확률이 높은 경로를 찾아서 확률, 이전 상태, 진입시간을 기록한다. 입력이 완료되면 기록된 정보를 이용하여 가장 확률이 높은 경로를 찾음으로써 최적의 어휘를 얻는다.

3.4 분산 어휘 인식 프레임워크

휴대용 단말기에서의 음성인식 한계를 극복하고 이동 중에서도 편리하게 이용할 수 있는 방법이 분산 어휘 인식 시스템이며 다양한 분야에 음성기술이 접목되고 있음에도 불구하고 휴대용 단말기에서는 CPU 용량의 한계로 구현의 어려움이 따른다.

컴퓨터와 무선 네트워크 통신을 이용해 휴대용 단말기와 휴대폰 등이 가지는 한계성을 극복하고 어휘 인식 기능을 구현할 수 있는 기술이 바로 분산 어휘 인식이다.

분산 어휘 인식은 이러한 문제를 해결한 기술로 휴대용 단말기에서는 음성신호의 특징을 인식하고 용량이 많이 소요되는 인식부문은 서버에서 처리하는 2원적 처리구조를 가지고 있다. 어휘 인식을 위한 다단계 프로세스 중 일부는 휴대용 단말기에서 처리하고 나머지 작업은 컴퓨팅 파워가 좋은 서버에서 수행한 다음 그 결과를 다시 휴대용 단말기가 받아 사용자에게 결과를 전달하는 일련의 작업이다. 그림 4는 분산 어휘 인식 시스템의 프레임 워크이며 이 프레임 워크에 의해 어휘인식 코드를 두 개로 분할해 휴대용 단말기와 서버에서 각각 처리할 수 있게 하고 입력 음성의 특징을 음소 모델을 통해 추출한 음소열을 인코딩을 거쳐 무선 네트워크를 통해 전달하면 서버에서 전송 받은 음소열을 디코딩을 거쳐 음향모델을 이용한 렉시컬 트리 탐색 알고리즘에 의해 단어를 찾아 그 결과를 휴대용 단말기로 재전송하여 음성으로 들려준다.

음성 특징의 데이터를 추출하는 과정을 수행한 이후 그 데이터를 압축하는 것까지 휴대용 단말기에서 수행한다. 압축된 음성 특징의 데이터를 무선 네트워크 방법인 DSR(Dynamic Source Routing)을 이용하여 서버로 전송하고 나머지 부분인 GMM 음소인식과 렉스컬 트리 탐색을 실행하게 하므로 인식의 정확성과 휴대용 단말기의 활용성이 높아지게 된다.

휴대용 단말기를 이용하여 분산어휘인식 시스템 구현을 위해, 클라이언트부인 휴대용 단말기에서는 음성입력과 전처리 부분을 담당하고, 서버부인 메인 컴퓨터에서는 어휘인식 엔진 부분을 구현한다.

DSR 방법은 각 패킷의 헤더에 패킷이 통과할 모든 노드들의 리스트가 포함된 형태의 소스 라우팅 방법으로 중간 노드들이 패킷 전달을 위해 최신 경로 정보를 유지할 필요가 없는 장점을 갖는다. 또한 주기적인 경로 광고 패킷이나 인접 노드 검사 패킷을 보낼 필요가 없다. 목적지까지의 경로 결정은 경로 발견 단계와 경로 유지 관리 단계로 진행되는데 먼저 소스 노드가 경로 요구 패킷을 방출하면 목적지까지

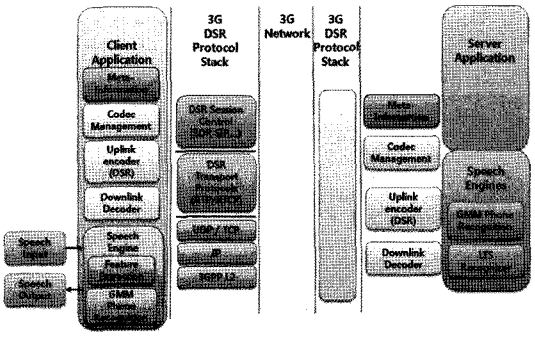


그림 4. 분산 어휘인식 시스템의 프레임워크

의 경로를 알고 있는 노드가 경로 응답 패킷으로 응답을 해서 경로를 찾아내고 만약 네트워크 토폴로지가 변해서 목적지까지 경로를 찾을 수 없는 경우, 즉 경로 탐색 타이머가 만료되면 소스 노드가 자신의 캐쉬에 있는 다른 경로로 시도하든지 다시 경로 발견 단계를 수행해서 경로를 찾아낸다[14-16].

4. 실험결과 및 분석

본 논문에서 제안한 어휘 인식 속도 향상에 대한 방법의 성능 검증을 위하여 인식 실험을 수행하였으며 시스템 환경은 표 1과 같다. 훈련과정과 실험환경과의 불일치 문제를 해결하기 위해 잡음처리는 워너 필터를 사용하였다. 어휘 목록은 회사명, 지역명, 지하철역명으로 구성하였다. 인식 실험에서는 실험에 참가한 화자가 어휘 중에서 임의로 100단어씩 발음하여 총 1500단어르 대상으로 실험을 수행하였다. 음성은 실내 환경과 잡음 환경에서 이동기기 등에 내장되어 있는 내장형 마이크로폰을 사용하여 16kHz Mono로 녹음 하였고, 16bit PCM 양자화를 사용하였다. 실험 음성은 실내 10명, 실외 5명 등 총 15명의 성인 남성이 참가하였다. 실내 환경은 50~55dB이고, 실외 환경은 70~75dB의 소음환경 하에서 실험하였다.

표 2는 일체형 구조와 분산형 구조의 인식 속도에 관한 실험 결과이다. 일체형 구조는 표 1의에서 제시한 서버에서 실험하였고 분산형은 표 1에서 제시한 클라이언트와 서버를 통해 실험하였다. 실험은 음성 입력이 끝난 상태에서 인식 결과가 나오기까지의 시간차를 이용해 측정하였으며 결과에 의하면 어휘수에 따라 일체형은 인식 속도가 증가하는 것을 보이는 반면 분산형은 같은 속도를 나타내고 있다. 따라서 인식 어휘가 늘어나도 제안한 분산형 구조에서는 인식 속도에는 영향을 미치지 않음을 확인할 수 있었다.

표 1. 시스템 환경

항목	Client	Server
모델명	IPAQ 112	Pentium 4
CPU	624MHz	2.4GHz
RAM	128M	512M
ROM	256M	
OS	Windows Mobile 6 Classic	Windows XP

표 2. 인식 속도

어휘수	인식 속도(Sec)	
	일체형	분산형
5만	1.15	1.58
15만	1.52	1.58
30만	1.83	1.58

표 3. 음소 인식

혼합 밀도	인식률	
	HMM	GMM
32	38.9	41.3
64	49.6	52.6
128	58.3	60.8
256	61.7	62.1
512	64.4	64.2

표 3에서는 음향모델을 이용한 음소인식기의 성능을 비교하였다. 두 방식 모두 모노폰 모델을 사용하였고, 혼합 밀도 중가에 비례하여 인식률이 증가하지 않고 일정 수준에 수렴되는 것을 볼 수 있다. 본 논문에서 사용한 GMM을 이용한 음소 인식기의 성능이 HMM에 비해 조금 우위에 있는 것을 확인 할 수 있었다. 이것은 대각 공분산 성분을 가지는 HMM 음소모델에 비해 모든 행렬의 공분산 성분을 가지는 GMM 모델이 변별력이 더 크기 때문이다.

표 4와 5는 일체형 인식 구조와 분산형 인식 구조를 실내 환경에서의 실험과 실외 환경에서의 실험을 나타낸다. 실내 환경은 50~55dB에서 실험 하였으며, 실외 환경은 70~75dB의 소음환경 하에서 실험

표 4. 실내 환경 인식률

어휘	인식률(%)	
	일체형	분산형
어휘 종속	97.51	98.01
어휘 독립	96.97	97.71

표 5. 실외 환경 인식률

어휘	인식률(%)	
	일체형	분산형
어휘 종속	91.01	91.01
어휘 독립	90.07	90.01

하였다. 결과에서 보는 것과 같이 시스템 성능 평가 결과 어휘 종속 인식률은 98.01%, 어휘 독립 인식률은 97.71%의 인식률을 나타냈으며 인식속도는 1.58초로 나타내었다. 이는 스탠드 얼론에서의 결과 값과 비교하였을 때 속도 면에서 0.3초 향상되었으며 인식률에서 1.1% 향상된 결과를 나타냈다.

5. 결 론

본 논문은 휴대용 단말기를 이용한 분산어휘인식 시스템에 관한 것으로 2원적 어휘인식모듈을 개발하고 휴대용 단말기 어휘인식에 관해 실험하였다.

분산 어휘인식 시스템은 휴대용 단말기에서는 간단한 음소인식을 수행하고 인식 결과를 서버로 전송하여 복잡한 단어 인식을 수행하는 방법이다. 어휘인식을 위한 음향학적 탐색은 휴대용 단말기에서 수행하고 보다 복잡한 언어적 탐색은 서버에서 처리하는 것으로 입력된 음성신호로부터 특징벡터를 추출하여 GMM을 이용한 음소인식을 수행하고, 인식된 음소 열을 서버로 전송하여 제안된 렉시컬 트리 탐색 알고리즘을 사용하여 언어적 탐색 단계에서 단어 인식을 수행한다. 제안한 분산 구조로 인하여 휴대용 단말기의 처리량을 줄임으로써 메모리 사용량이 현저히 줄일 수 있었으며 인식 속도와 인식률에서 일체형보다 나은 결과를 얻을 수 있었다. 시스템 성능 평가 결과 어휘 종속 인식률은 98.01%, 어휘 독립 인식률은 97.71%의 인식률을 나타냈으며 인식속도는 1.58초로 나타내었다. 이는 스탠드 얼론에서의 결과 값과 비교하였을 때 속도 면에서 0.3초 향상되었으며 인식률에서 1.1% 향상된 결과였다.

따라서 분산 어휘 인식시스템은 기존의 인식 시스템에 비하여 휴대용 단말기를 이용하므로 휴대성과 장소의 구애를 받지 않고, 기존 인식 시스템의 높은 인식률을 그대로 적용할 수 있어 매우 효율적이라 할 수 있다.

참 고 문 헌

[1] 오지영, 김윤중, 고유정, "모바일 환경에서 인증과 음성인식을 위한 웹 서비스 구현," 한국멀티미디어학회 논문지, Vol.8, No.2, pp. 225-232, 2005.

[2] 김기백, 최종호, "음성인식 기반 컨텐츠 네비게이션 시스템," 한국컴퓨터정보학회 논문지, Vol. 15, No.1, pp. 99-102, 2007.

[3] 김승희, 황규웅, 전형배, 정훈, 박준, "분산어휘 인식을 위한 내장형 고속 및 경량 음소인식기 개발," 한국정보처리학회, 춘계학술발표대회, pp. 395-396, 2007.

[4] 윤경섭, "휴대용 단말기를 위한 실시간 무선 영상 음성 전송 기술," 한국컴퓨터정보학회 논문지, Vol.14, No.4, pp. 111-117, 2009.

[5] 방기덕, 강철호, "차량용 항법장치에서의 관심지 인식을 위한 다단계 음성 처리 시스템," 한국멀티미디어학회 논문지, Vol.12, No.1, pp. 16-25, 2009.

[6] M. F. Gales, "Model-based techniques for noisie robust speech recognition," Ph. D. dissertation, University of Cambridge, Sept, 1995.

[7] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, Prentice-Hall, 2000.

[8] David Pearce, "An overview of the ETSI standards activities for Distributed Speech Recognition Front-ends," *The Speech Applications Conference*, May 22-24, 2000.

[9] N. Srinivasamurthy, A. Ortega and S. Narayanan, "Efficient Scalable Encoding for Distributed Speech Recognition," Department of Electrical Engineering-Systems, Signal and Image Processing Institute, Integrated media Systems Center, March 28, 2004.

[10] A. S. Manos and V. W. Zue, "A study on out-of-vocabulary word modeling for a segment-based keyword spotting system," Master Thesis, MIT, 1996.

[11] T. Jitsuhiro, S. Takatoshi, and K. Aikawa, "Rejection of out-of-vocabulary words using phoneme confidence likelihood," ICASSP, pp. 217-220, 1998.

[12] Kris Demuynck, Tom Laureys, Dirk van Compernolle, and Hugo van Hamme, "FLavor: a flexible architecture for LVCSR," In EUROSPEECH-2003, pp. 1973-1976, 2003.

[13] L. Rabiner and B. H. Juang, *Fundamentals of*

Speech Recognition, Prentice-Hall, 1993.

[14] 곽운용, 오훈, "무선 이동 애드 혹 네트워크를 위한 동적 그룹 소스 라우팅 프로토콜," 한국통신학회 논문지, Vol.33, No.11A, pp. 1034-1042, 2008.

[15] 하은용, "이동 애드혹 네트워크에서 DSR 프로토콜을 위한 경로 축소 방법," 한국정보과학회 논문지, Vol.34, No.6, pp. 475-482, 2007.

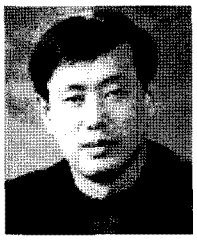
[16] David B. Jhonson, David A. Maltz and Yih-Chun Hu. "The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks for IPv4," RFC 4728, Feb. 2007.

안 찬 식

2002년 광운대학교 컴퓨터공학과 석사

2004년 광운대학교 컴퓨터공학과 박사수료

관심분야 : 음성인식, 분산처리, 음성/음향 신호처리



오 상 엽

1991년 광운대학교 전자계산학과 석사

1999년 광운대학교 전자계산학과 박사

1993년~현재 경원대학교 IT대학 컴퓨터소프트웨어 교수

관심분야 : 소프트웨어공학, 버전관리, 소프트웨어 재사용, 형상관리, 객체지향, 음성인식, 분산처리, 음성/음향 신호처리