PDA 상에서 음성명령어를 구현하기 위한 음성인식기의 설계

*곽상훈, 김 철, 최승호
동신대학교 정보통신공학과

The design of Speech Recognizer to Implement the Voice Command on the PDA

*Sang-Hun Kwak, Cheol Kim, Seung-Ho Choi
Dept. of Information and Communication Eng., Dongshin University
e-mail: shchoi@white.dongshinu.ac.kr

요 약

본 논문에서는 PDA상에서 음성으로 명령어를 제어하기 위해 Window CE 3.0 환경에서 음성인식기를 설계하였다. 전처리과정에서 26차 특징파라미터를 추출하고, HTK를 통해 학습하였다. 트라이폰 기반의 가변어휘 음성인식기를 설계하였으며, PDA의 응용프로그램은 Embedded Visual C++언어를 사용하여 22개의 음성명령어를 제어하도록 하였다. 그 결과 PDA상에서 92%의 인식률이 나타났으며 이것은 음성인식이 모바일 환경에서도 접근이 가능함을 알 수 있었다.

1. 서론

최근 PDA는 기존의 통신 시스템의 가장 큰 문제점인 단말기의 이동성과, 전송속도의 한계를 극복하기 위한 적합한 솔루션으로 자라잡고 있다[1]. 이러한 PDA에 최근 음성인식기술이 적용되어 이동 중에도 쉽게 단말기를 제어할 수 있도록 데이터를 더욱 빠르고 편리하게 전송하는 HCI기술이 대두되고 있다. 하지만 이러한 기술은 데이터의 보안문제, 과다한 무선접속비용, 상이한 표준으로 인한 호환성 결여 등의 문제점을 갖고있으며 특히 음성인식기술을 도입함에 있어 잡음을 포함한 모바일 환경이 커다란 문제로 인식되고 있다[2].

따라서 본 논문에서는 PDA 상에서 음성으로 명령어를 제어하기 위해 가변 어휘 인식기로 제한된 범위에서 인식기를 구현하였고, 타당성을 입증하기 위해 Stand-alone 모델, Client-Server 모델, PDA를 적용하여 비교실험 하였다.

2. 전처리와 특징추출

(1) 음성구간 검출

음성구간 검출은 전체 인식 처리 속도에 영향을 주지 않기 위해 실시간에 가까운 처리속도가 요구된다. 따라서 입력신호의 매 구간에서 구한 에너지 값과 주파수 특성을 고려한 영교차율을 구한뒤 미리 계산된 통계에 의해 결정된 임계값과 비교하고 음성과 묵음구간을 판별하였다.

다음은 평균에너지와 ZCR를 나타낸 것이다.

$$E_s = \frac{1}{N}\sum_{n=1}^{N} s(n)^2 \qquad (1)$$

여기서, N은 샘플의 개수를 의미한다.

$$sign[s(n)] * sign[s(n+1)] < 0 \; ; \; ZCR++ \qquad (2)$$

(2) 특징파라미터 추출

음성인식에 필요한 음가정보를 효과적으로 나타내는 특징 파라미터는 인간의 청각 특성을 고려한 MFCC를 사용하였다.

캡스트럼은 음성 신호로부터 성도에 대한 정보를 추출한 것으로 단구간 스펙트럼에 대한 로그 스케일의 크기를 역퓨리에 변환한 것이며 다음 식과 같이 정의 될 수 있다.

$$c'(n) = \sum_{k=1}^{K} (\log S'_k) \cos[n(k-\frac{1}{2})\frac{\pi}{K}] \qquad (3)$$

여기서, n은 캡스트럼의 차수이고, k는 필터의 차수를 나타낸다[3].

3. 음성인식기 설계

본 논문에서는 가변어휘 인식기를 설계하였으며 비교실험을 위해 연속HMM 음성인식기를 사용하였다.

(1) 연속 HMM을 이용한 음성인식기

연속HMM인식시스템은 HMM 알고리즘의 초기화, 학습과정 그리고 인식과정 등으로 구성할 수 있다.

학습과정은 HMM의 전후향 알고리즘과 바움-웰츠 알고리즘을 이용해서 음소를 모델링하고 각 음소 상태에 대한 평균과 분산 값, 상태천이확률 등을 얻게 된다.

인식과정은 학습에서와 같은 전처리 과정을 거쳐 파라미터를 추출한 후 가우시안 혼합모델에 의한 관측확률분포를 측정하고 전향 알고리즘을 적용하여 상태변이 확률을 계산한다. 이때 구해진 확률에 의해 최적의 확률값을 갖는 단어를 인식단어로 결정한다.

그림 1는 HMM을 이용한 인식과정을 나타낸 것이다.

(2) 가변어휘를 이용한 음성인식기

가변어휘 음성인식을 하기 위해서는 첫째, 텍스트가 음운 변동과정을 수행하고, 둘째, 학습된 데이터를 참고로 결정트리 기반 상태 공유 알고리즘을 이용하고, 셋째, 각 트라이폰에 대한 상태를 구하고, 넷째, 이를 다시 인식에 사용할 수 있는 어휘사전구조로 재구성하고, 다섯째, 탐색 공간이 넓으면 넓을수록 탐색의 정확도가 증가되기 때문에 빔 탐색 방법을 사용한다.

가변어휘에서 정의된 질의어는 입력 트라이폰에 대해 결정트리를 구성하고 이 트리는 각 노드에 하나의 질의어를 갖는 이진 트리로 좌우 대칭구조를 가진다[4].

결정트리 방법으로 하향식을 사용하며 이것의 특징은 학습 중에 나타나지 않는 음소에 대해서도 모델링할 수 있는 확장성을 지녔기 때문이다[6].

그림 2는 가변어휘 인식과정을 나타낸 것이다.

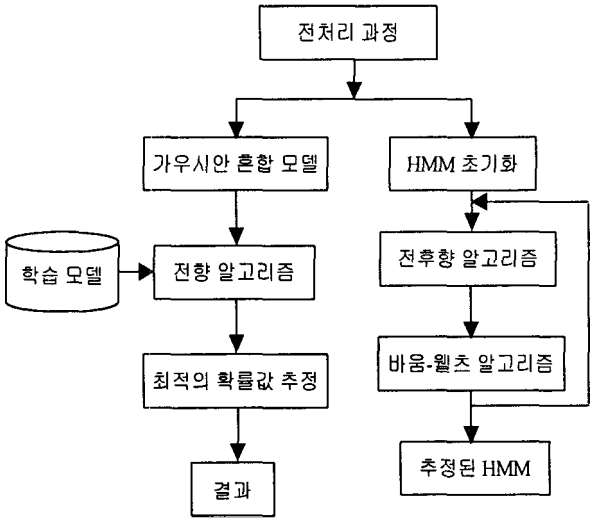

그림 1. HMM을 이용한 인식과정

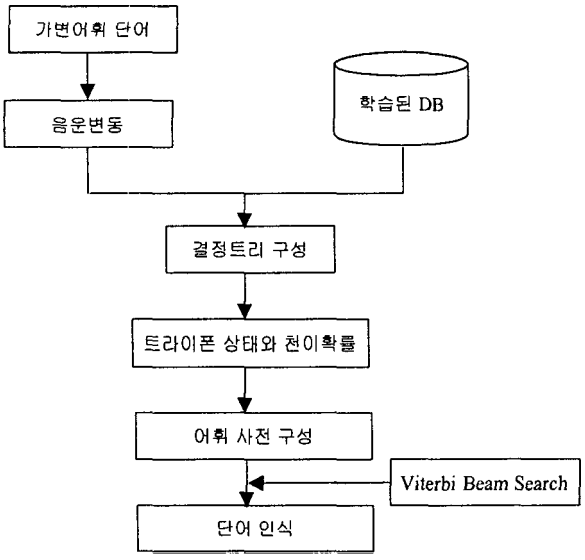

그림 2. 가변어휘 음성인식과정

(3) PDA에서 구현된 응용 프로그램

PDA에 사용되는 음성 인식 프로그램은 Microsoft의 Window CE 개발 도구인 Embedded Visual C++을 사용하여 개발하였으며 다음의 소스는 PDA의 화면에 출력되는 소스의 일부분을 나타낸 것이다.

C/S모델과 PDA를 구분하여 인식실험 하기 위해 사용자 인터페이스를 각각 그림3과 4에 나타내었다.

```
BOOL CRecogLiveProDlg::OnInitDialog()
{
        CDialog::OnInitDialog();
        SetIcon(m_hIcon, TRUE);
SetIcon(m_hIcon, FALSE);
CenterWindow(GetDesktopWindow());
OnInitSound(AfxGetApp()->m_pMainWnd->m_hWnd,
32, 2000);
OnSetSoundParam(1, 8000, 16);
ctrlWnd = GetDlgItem(IDC_SIGNAL)->m_hWnd;
OnSetRecogParam(8000, 12, 22, 26);
OnInitRecog();
OnCalib();
GetDlgItem(IDC_CALIB)->EnableWindow(FALSE);
GetDlgItem(IDC_RECOG_START)->EnableWindow(T
RUE);
return TRUE;
}
```

그림 3. PDA에서 구현된 데모 화면

그림 4는 Client-Server모델이 웹상에서 구현된 사용자 인터페이스를 나타낸 것이다[5].

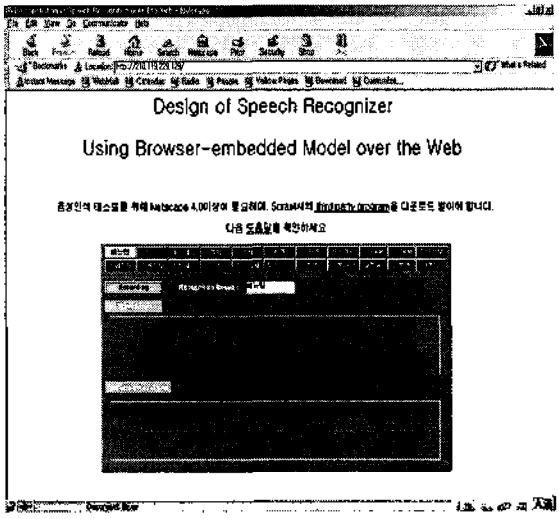

그림 4. 웹상에서 구현된 Client-Server모델의 사용자 인터페이스

4. 실험 및 고찰

(1) 실험환경

본 논문에서 사용하는 PDA는 컴팩사의 iPAQ H3660 모델을 사용하였고 빠른 데이터 전송과 음성처리를 위해 기본 메모리 64M에 128M의 CF 메모리를 추가하여 실험하였다.

Client-Server 모델을 실험하기 위해서 서버 컴퓨터에서 O/S는 Window 2000을 사용하였고 JDK는 1.3베타 버전, C 컴파일러는 Visual Studio 97, 오디오 클래스는 SoundBite 1.0을 사용하였다. 이에 반해 클라이언트 컴퓨터에서는 Window 98, 웹브라우저는 넷스케이프 6.0을 사용하였으며 오디오 클래스는 마찬가지로 SoundBite 1.0을 사용하였다.

(2) 인식실험

데이터베이스로 학습과 인식의 음성데이터는 20대 남성화자 70명을 대상으로 실험하였으며, 그중 52명은 학습, 18명은 인식 실험에 사용하였다.

Stand-alone과 Client-Server 모델에 대해 학습에 참가한 22(단어/명)×18명=396개 단어를 TEST1, 학습에 참가하지 않은 화자 22(단어/명)×5×2회=220개 단어를 TEST2로 사용하여 각각 인식실험을 하였다.

표 1. 모델에 대한 인식률 비교

Model	인식기반	인식수/실험횟수	인식률(%)
Stand-alone	CHMM	392/396(TEST1)	99%
	가변어휘	376/396(TEST1)	95%
Client-Server	CHMM	202/220(TEST2)	90.9%
	가변어휘	187/220(TEST2)	87%
PDA	CHMM	213/220(TEST2)	97%
	가변어휘	202/220(TEST2)	92%

표 1에서는 모델에 관계없이 TEST1과 2에서 가변어휘 성능이 CHMM보다 인식률이 낮게 나타났다, 이를 분석해보면 인식 대상단어가 너무 적어 다양한 음소모델을 포함하지 못한다고 판단된다.

Client-Server 모델에서는 TEST2를 대상으로 웹상에서 구현되었기 때문에 인식률이 저하되었다.

PDA상에서는 가변어휘가 Stand-alone 모델보다 약 3%의 성능저하를 가져왔는데 이는 PDA 자체적인 하드웨어 특성상 발생하는 신호 왜곡이 가장 큰 원인으로 볼 수 있다.

5. 결론

본 논문에서는 PDA상에서 명령어를 제어하기 위해 가변어휘를 이용한 음성인식기를 설계하였으며 Window CE 3.0에서 Embedded Visual C++언어로 응용프로그램을 연구하였다. 음성인식실험은 CHMM과 가변어휘 인식기를 사용하였으며 Stand-alone 모델, Client-Server 모델, PDA에 적용하여 실험하였다. 그 결과 PDA에 음성인식기의 내장은 가능하지만 데이터의 용량과 처리 능력이 고려되어야하며 인식프로그램을 최적화하고 경량화 하는데 많은 시간이 소요되었다.

향후에는 PDA에서의 음성인식기를 구현할 때 나타나는 문제점들을 보완하여 PDA를 클라이언트로, PC를 서버로 하는 모바일 컴퓨팅에 적합한 C/S 모델의 가변어휘 음성인식기를 자바로 구현할 계획이다.

참고문헌

[1] 배찬권, "세계 PDA 시장 전망," KISDI IT FOCUS, 1월호, 2001.

[2] 이상오, " Palm의 경영위기 배경과 PDA 시장의 전개방향," 정보통신산업연구실, 2001.

[3] Claudio Becchetti, Lucio Prina Ricotti, "Speech Recognition," WILEY&SONS, 1999.

[4] Steve Young, The HTKBook(for version 2.2), Entropic LTD., 1999.

[5] 최광국, 김철, 최승호, 김진영, "자바를 이용한 음성인식시스템에 관한 연구," 한국음향학회지, 제 19권, 제 6호, 2000.

[6] 서봉수, "가변 어휘 음성 인식기 구현 및 탐색 시간 단축 알고리듬 비교" 석사학위논문., 전남대학교 전자공학과, 2001.

※ 본 논문은 2001년도 정보통신부 대학기초 연구 자원사업으로 수행된 연구 결과물입니다.