

가변어휘 인식기를 이용한 PDA상에서의 음성제어 구현

곽상훈(동신대), 최승호(동신대), 신도성(전남대), 김진영(전남대)

<차 례>

- | | |
|--------------------------|------------------|
| 1. 머리말 | 4. SDK의 설계 및 구현 |
| 2. PDA의 구조 및 사양 | 4.1. SDK |
| 2.1. PDA의 운영체제 | 4.2. 응용프로그램 알고리즘 |
| 2.2. Windows CE | 5. 실험 및 고찰 |
| 2.3. 시스템의 기본구성 | 5.1. 실험환경 |
| 3. 가변어휘 인식기 설계 | 5.2. 인식실험 및 고찰 |
| 3.1. 트라이폰 학습과정 | 6. 맺음말 |
| 3.2. 결정 트리 기반 상태 공유 알고리즘 | |

<Abstract>

Implementation of Voice Control on PDA using the Text Independent Vocabulary Recognizer

Sang Hun Kwak, Do Sung Shin, Seung Ho Choi, Jin Young Kim

The technology of speech recognition has a wide field of application. The range of such technology is spreading into mobile computing having the large amount of movement for communication equipments at the present time. Particularly, recognition in internet environment is rapidly moving into mobile environment. Because of these environments, users want the faster speed of data transmission and the lighter portable equipment for data access. That is PDA(Personal Digital Assistant).

Therefore, we designed a triphone-based text independent vocabulary recognizer for the implementation of speech control in this paper. The text independent vocabulary recognizer is based on the state joint algorithm with decision trees

* 주제어: PDA, Windows CE, 가변어휘, 트라이폰, 결정트리, SDK

1. 머리말

인류의 가장 확실한 정보통신 수단인 음성은 다른 인터페이스보다 뛰어난 장점들을 가지고 있으며 최근에는 컴퓨터의 발달과 정보통신 서비스의 발전으로 인해 음성처리에 대한 연구가 활발하게 진행되고 있다.

음성인식은 음성을 기계가 인식할 수 있도록 하는 기술을 말하는데 이러한 음성인식 기술은 동적인 프로그램을 이용하여 입력 음성과 기준이 되는 학습 음성을 비교하여 최소 거리를 갖는 음성을 입력으로 인식하는 DTW, 음성을 상태천이 확률과 각 상태에 대한 관측 확률을 갖는 마코프 과정으로 가정한 후, 학습하고 학습된 확률 정보를 이용하여 인식하는 은닉 마코프 모델 그리고 인간의 신경세포 구조와 기능을 모델화하여 사용하는 신경회로망 인식 등이 있다.

최근 음성인식 기술의 응용분야는 시공간을 초월해 빠른 속도로 확산되고 있다. 또한 정보통신 기기의 이동성이 이슈로 부각되고 있는 현재의 상황에서 음성인식의 기술은 모바일 시장을 타깃으로 그 영역을 넓혀가고 있다. 특히 인터넷 시장의 폭발적인 성장으로 사용자들의 인터넷 환경에 대한 인식이 모바일 환경으로 발전하게 되었고 기술발전도 자연스럽게 따라가게 되었다 [3].

모바일 컴퓨팅의 솔루션은 여러 분야로 발전되고 있는데 빨라진 데이터 전송속도, 작고 가벼운 휴대형 장치로 데이터에 더욱 편리하게 접근하고자하는 고객의 욕구를 만족시키는 것이 바로 PDA(Personal Digital Assistant)이다. 현재 휴대의 극대화를 위해 더 이상 키보드만으로 PDA를 컨트롤하기가 힘들어짐에 따라 음성인식 기술은 PDA와 같은 소형단말기에 필요한 기술이다.

2. PDA의 구조 및 사양

PDA는 CPU, 펜이나 터치스크린 입력을 통한 데이터 입력장치 그리고 통신장치 세 부분으로 구성된다. 그 중 데이터 입력방식은 크게 펜 입력, 터치스크린 입력, 바코드 스캐너 입력 그리고 음성입력으로 나뉜다. PDA에 장착된 유무선 통신기능을 이용하면 시간과 장소에 관계없이 다른 정보기기와 문자, 음성, 영상 등의 정보를 송수신하고, 다양한 종류의 서비스를 제공받을 수 있다[1].

2.1. PDA의 운영체제

PDA에 있어 운영체제(OS : Operating System)는 CPU뿐만 아니라 메모리, LCD, 디지털타이저, 통신장치 등 하드웨어의 기본적인 인터페이스 운영방안을 제공한다. 특히 PDA가 휴대성을 강조한 초소형 제품인 만큼 PDA에 사용되는 OS 역시 크기

가 작고 사용자 인터페이스도 편리하며, 응용소프트웨어 개발이 용이해야 한다. 현재 대표적인 PDA의 OS는 Microsoft의 Windows CE, Palm Pilot의 OS인 Palm OS, Symbian의 EPOC 등이 있으며, 본 논문에 사용된 OS는 Windows CE 3.0을 사용하였다.

2.2. Windows CE

Windows CE는 마이크로소프트사가 지능적인 작동을 요구하는 가전기기, 제어장치 및 휴대용 장비를 운영하기 위해 개발된 작고 이식성이 뛰어난 OS로서 Windows 95/98, Windows NT와 비슷한 운영체제이다. 현재 Windows CE는 소형 모바일 컴퓨터(HPC, Wallet PC, Palm PC), 무선 통신 기기(디지털 호출기, 셀룰러 폰), 차세대 오락/멀티미디어 기기, 특수 목적용 인터넷 접속기(인터넷 TV, 디지털 셋탑 박스) 그리고 정보화 기기 등 다양한 범위에서 사용하고 휴대가 가능한 새로운 시스템으로 평가받고 있다. 본 논문에서 사용되어진 Windows CE 3.0은 더욱 향상된 Windows 호환성을 제공하며, 보다 커진 저장소 및 파일 처리 기능, 프로세스간 통신(IPC), 네트워킹 지원 기능 등을 사용함으로써 Windows NT와 Windows 2000을 기본으로 하는 데스크톱 환경과 쉽게 상호 작용하게 된다. 따라서 응용 프로그램 개발자는 마이크로소프트의 ActiveX 컨트롤과 메시지 대기열(MSMQ), 구성 요소 개체 모델(COM) 인터페이스, ATL(Active Template Library), MFC(Microsoft Foundation Classes)라이브러리 등에 대해 Window CE 3.0의 지원 기능을 활용할 수 있다[2].

2.3. 시스템의 기본 구성

시스템의 기본구성은 포켓 PC와 데스크탑 PC를 USB-Sink 케이블로 서로 연결하고 Microsoft ActiveSync를 이용하여 포켓 PC와 데스크탑 PC를 서로 동기화 한다. 이렇게 데스크탑과 포켓 PC를 동기화 함으로써 데이터를 서로 비교해서 변경된 내용들을 맞추며, 양쪽의 데이터들을 최신의 내용으로 업데이트 해준다.

ActiveSync은 메인 데스크톱 PC와 포켓 PC 사이에 동기화 서비스를 제공함으로써 사용자 자신의 개인 정보를 포켓 PC를 이용해 어느 곳이나 가지고 다닐 수 있다. 또한 데스크톱과 포켓 PC 사이에 개인 정보를 전달할 뿐만 아니라 코드를 개발하고 연결된 장치에서 해당 코드를 자동으로 실행할 수 있도록 만들어준다. ActiveSync의 특징은 다음과 같다. ActiveSync가 지원되는 Windows 버전에서는 전화 접속 네트워킹을 설치할 필요가 없으며 USB, 적외선, 전화 접속 및 인터넷 등의 연결뿐만 아니라 직렬 포트를 통해서도 실행되고 PC와 다른 장치의 통신에 사용되는 포트를 자동으로 감지하여 구성할 수 있다.

3. 가변어휘 인식기 설계

일반적인 음성인식기는 인식 대상 단어가 고정되어 있어서 대상 단어가 변경될 때 인식기의 데이터베이스를 재구축 해야 하는 단점이 있으나 가변어휘 인식기는 대상 단어의 변경이 손쉽고 사용자가 원하는 대상 단어를 자유롭게 입력할 수 있다는 장점을 가지고 있다 [4].

3.1. 트라이폰 학습과정

본 논문에서는 자동차 항법장치에 쓰일 수 있는 지명, 도로명, 상호 그리고 건물 등의 단어들을 수집하여 이를 음운 변동을 거쳐 트라이폰(Tri-phon) 구조로 바꾸었다. 트라이폰이란 어떤 음소의 좌우에 위치하는 음소는 가운데 위치한 음소에 아주 많은 영향을 주는데 이런 영향까지 고려한 인식단위를 말하는데 일반적으로 트라이폰은 그 개수가 많아서 충분한 학습이 어렵지만 조음화 현상을 모델링하는 인식단위로 큰 장점을 지닌다.

트라이폰 종류를 포함하는 단어 선정 순서는 다음과 같다. 우선 음운 변동을 거친 단어들을 영문 발음으로 바꾸고 이를 트라이폰 구조로 바꾼다. 트라이폰 구조로 바꾼 단어들에서 트라이폰 종류를 탐색하고 이를 빈도에 따라 오름차순으로 정리하고, 오름차순으로 정리된 트라이폰에 대해 순차적으로 모든 단어를 검색하여 검색할 트라이폰을 포함하는 단어를 선택한다. 선택한 단어와 이전 선택된 단어를 비교하여 새로운 트라이폰을 더 많이 가지고 있는 단어를 선택하며 모든 단어에 대해 탐색이 끝나면 선택한 단어를 구성하는 트라이폰과 선택 단어의 flag를 1로 하여 다음 검색에서 제외시킨다. 이 모든 트라이폰에 대해 탐색이 끝날 때까지 다시 반복하여 실행한다 [6].

녹음 환경은 잡음이 없는 조용한 사무실에서 콘덴서 마이크를 사용하였으며, 녹음 데이터는 16 kHz, 16 bit 샘플링이나 실제 학습과정에서는 8 kHz, 16 bit로 다운 샘플링해서 사용하였다. 녹음된 음성데이터는 HTK를 통해 학습하였다 [5]. 이를 위해 먼저 음운 변동과정과 한국어 발음을 영어 발음으로 바꾸는 과정을 함께 한다. 한국어 변이음 유니트를 이용해서 바뀐 단어와 그 단어에 해당하는 음성데이터를 가지고 학습과정을 거치는데 이는 HMM의 전-후향 알고리즘과 바움-웰츠 알고리즘을 이용해서 음소 모델링을 하여 각 음소 상태에 대한 평균과 분산 그리고 상태 천이 확률 등을 추출하게 된다 [7].

3.2. 결정 트리 기반 상태 공유 알고리즘

대어휘 음성인식 시스템에서는 연속분포 은닉 마르코프 모델(CHMM)을 주로 사

용하지만 이는 많은 숫자 모델의 파라미터들로 인한 데이터 부족을 초래하며 모델별 학습데이터의 수가 균일하지 않은 경우가 일반적이다. 이러한 단점을 보완하는 대표적인 접근 방법 중 본 논문에서는 이미 다른 방법보다 우수한 하향식 상태기반 공유를 사용하였고, 이 방법은 모든 문맥들에 대한 문맥종속 모델을 찾을 수 있으므로 학습용 데이터에는 나타나지 않았던 문맥들을 모델링 할 수 있으며, 질의어의 집합을 생성할 때 전문자의 지식이 사용되고, 트리생성 과정에 적당한 제약을 둬으로서 각 모델별로 충분한 데이터를 확보하면서 동시에 상태들의 개수를 적정 수준으로 유지할 수 있다는 장점을 가진다 [7][9].

3.2.1. 질의어

질의어란 학습과정에서 각 트라이폰들이 가지고 있는 상태를 공유하기 위하여 각각의 트라이폰들의 좌, 우측의 음소특성을 고려하여 구분 지어 놓은 것을 말하며 이 질의어가 어떻게 정의되느냐에 따라 트라이폰 상태 공유 정도는 달라지게 되고 이에 따라 인식률에도 영향을 미치게 된다.

표 1은 음성 질의어 집합을 구성하기 위해 자음과 모음을 음운학적인 특징에 따라 분류한 것이다. 또한 HTK를 통해 학습을 하게 되면 각 트라이폰에 대한 상태들을 공유하게 되는데, 이때 이러한 상태 공유를 결정짓게 되는 중요한 질의어 셋이 된다. 질의어 종류의 의미는 표 2에 나타내었다.

<표 1> 질의어 집합 구성을 위한 음성 분류

음운특징	음성 질의어 그룹
유, 무성음	Voiced_All, Voiced_Top, Unvoiced_CONs, Voiced_Cons, Un_Voiced_Top
혀의 위치	Top, High, Medium, Low, Front_Top, Central_Top, Back_Top
발성 위치	C_Front, C_Central, C_Back, V_Front, V_Central, V_Back, Front, Central, Back, Anterior, NonAnterior
비음, 파찰음, 파열음, 마찰음, 유음, 경음, 격음	Nasal, Central_Fric, Back_Fric, Affricate, NonAffricate, Plosive, Fricative, Tense, Fortis
모음, 자음	Vowel, AVowel, UVowel, EVowel, IVowel, OVowel, Diphthong
발성 방법, 입 모양	Coronal, NonCoronal, Rounded, UnRounded

<표 2> 질의어 종류와 의미

Top : 위쪽에서 발성	Front : 앞에서 발성
Nasal : 비음	Central : 가운데서 발성
Fricative : 마찰음	Back : 뒤에서 발성
Liquide : 유음	Fortis : 격음
C_Front : 앞에서 발성되는 자음	Lenis : 연음
C_Central : 가운데서 발성되는 자음	Coronal : 혀끝으로 발성
C_Back : 뒤에서 발성되는 자음	NonCoronal : 혀끝으로 발성 안되는 것
V_Front : 앞에서 발성되는 모음	Anterior : 앞쪽에서 발성
V_Central : 가운데서 발성되는 모음	NonAnterior : 앞쪽에서 발성 안되는 것
V_Back : 뒤에서 발성되는 모음	Voiced_Top : 유성음 중 혀의 위치가 위쪽 인 것
Continuent : 발성이 지속되는 것	UnVoiced_Top : 무성음 중 혀의 위치위쪽 인 것
NonContinuent : 발성이 지속 안되는 것	Front_Top : 혀의 위치가 앞쪽 위
Trident : 세 갈래로 나뉘어서 발성	Central_Top : 혀의 위치가 중간 위
Glide : 반모음	Back_Top : 혀의 위치가 뒤쪽 위
UnVoiced_Cond : 무성 자음	Tense : 경음
Voiced_Cond : 유성 자음	
UnVoiced_All : 무성음	
Diphthong : 이중 모음	

3.2.2. 질의어를 이용한 결정 트리 구성 및 상태 공유

정의된 질의어를 이용하여 입력 트라이폰에 대한 결정트리를 구성하며 이러한 트리는 각 노드에 하나의 질의어를 갖는 이진트리로 좌우 대칭구조를 가진다. 질의어 집합은 그림 1과 같이 표현이 되며 결정트리는 다음의 몇 가지 조건을 만족해야 한다.

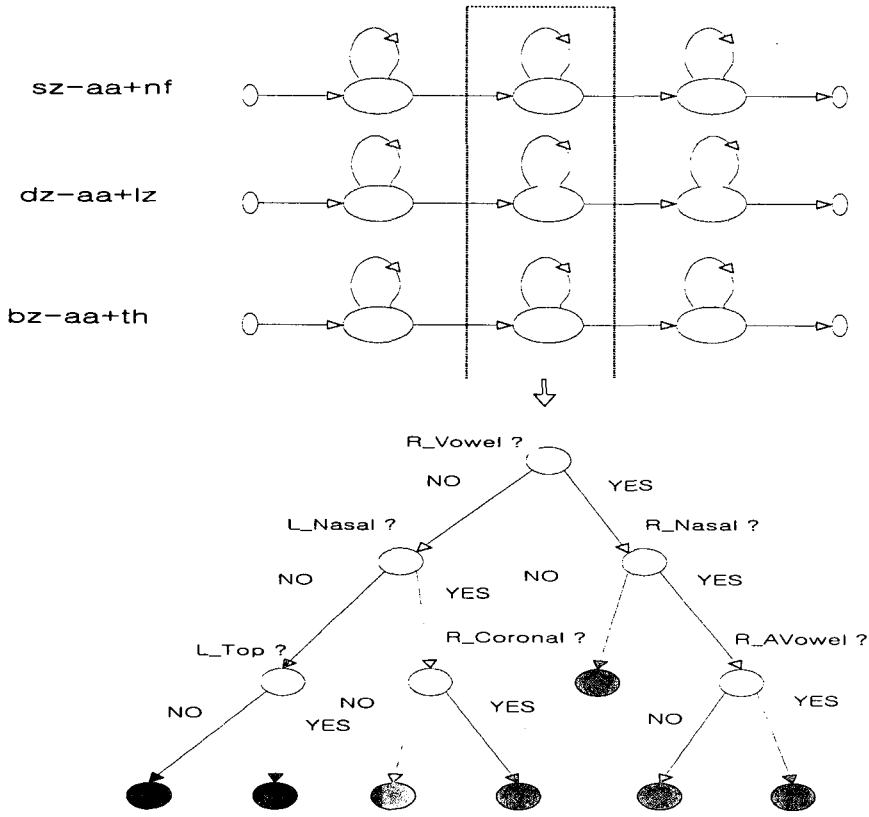
첫째, 최종 모델의 파라미터를 정확하고 강인하게 추정하기 위하여 각 단말 노드는 일정 개수 이상의 예를 반드시 가지도록 해야 한다.

둘째, 노드 분할을 위하여 질의어의 집합을 사용함으로써 노드 분할 과정을 제어함과 동시에 이 과정에 전문가의 지식이 사용될 수 있도록 한다.

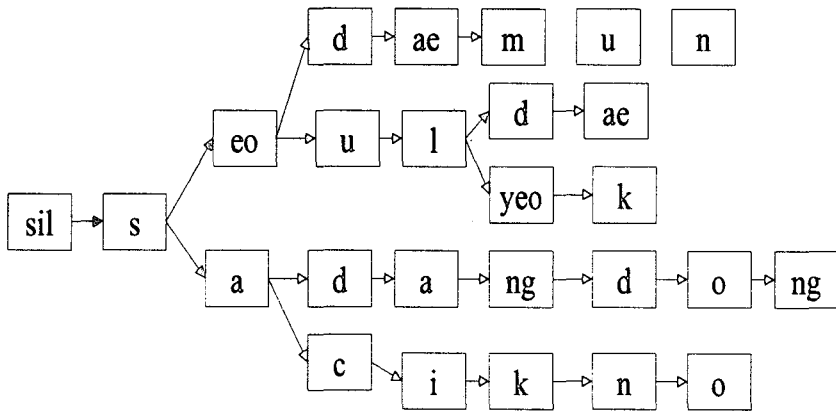
셋째, HMM은 여러 가지 단말노드들을 정확하게 표현 할 수 있어야 한다. 이를 위해서 혼합 가우시안 확률 분포가 각 단말 노드별로 훈련 데이터를 정확히 모델링 할 수 있어야 한다.

만들어진 트리를 이용해서 인식 시스템에서 인식 하고자하는 단어집합들을 입력으로 하면 인식기에 쓰일 수 있도록 각 트라이폰에 대한 상태와 천이 확률들로 정의된 파일을 만들어준다.

본 논문에서는 이렇게 구해진 각 입력 단어들의 트라이폰에 대한 상태들을 이용하여 그림 2와 같이 어휘사전(Lexicon) 구조를 만들게 된다.



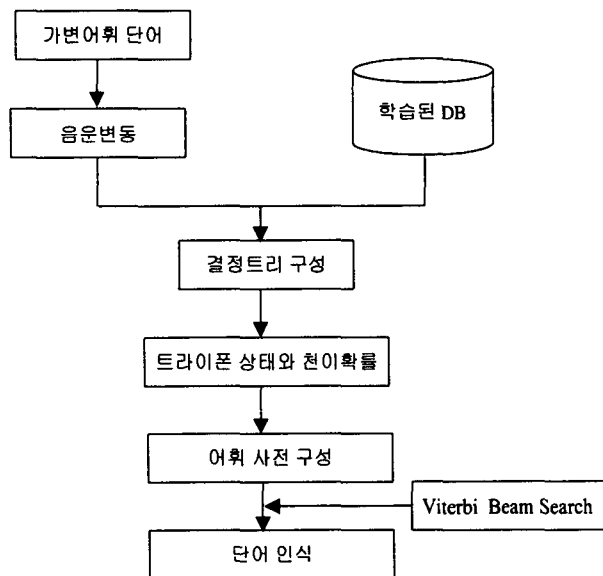
<그림 1> 결정트리의 예



<그림 2> Lexicon 구조

3.2.3. 가변어휘 인식과정

가변어휘 음성인식은 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. 그림 3은 가변어휘 인식 과정을 나타낸 것이다.



<그림 3> 가변어휘 인식과정

4. SDK의 설계 및 구현

4.1. SDK(Software Development Kit)

Windows CE 기반의 응용 프로그램을 개발하기 위해 다양한 종류의 개발 도구를 선택할 수 있는데 사용자는 C 또는 C++을 사용하여 Win32R API를 작성하거나, C++클래스 프레임 워크를 원하는 사용자는 MFC 라이브러리를 사용할 수 있다[7].

현재 가장 대표적인 것은 Platform Builder 3.0 과 embedded Visual Tools 3.0과 같은 두 가지 기본 개발 도구가 있다. 먼저 Platform Builder는 Windows CE 자체의

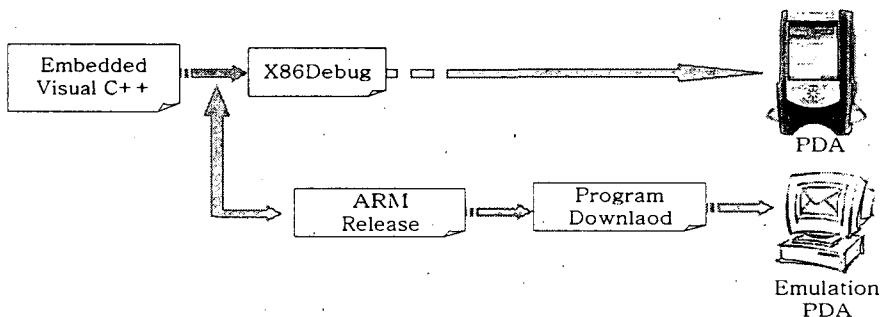
사용자 정의 설정을 만들기 위해 사용하는 도구이다. 다음으로 embedded Visual Tools 3.0은 데스크탑 개발환경에서 강력해진 장치, 포켓 PC 그리고 PC 등을 포함한 윈도우에 대한 응용 프로그램이나 시스템 구성요소를 만드는데 사용된다. embedded Visual Tools 3.0에는 embedded Visual Basic과 embedded Visual C++이 포함되어 있다. Visual basic을 자주 사용하는 사용자의 경우 embedded Visual Basic을 사용하여 Windows CE를 대상으로 하는 Visual Basic 기반의 응용 프로그램을 만들 수 있다 [8].

표 3은 embedded Visual Tools 3.0을 사용하기 위한 기본적인 시스템 요구사항이다.

<표 3> Embedded Visual Tools 3.0을 위한 시스템 요구사항

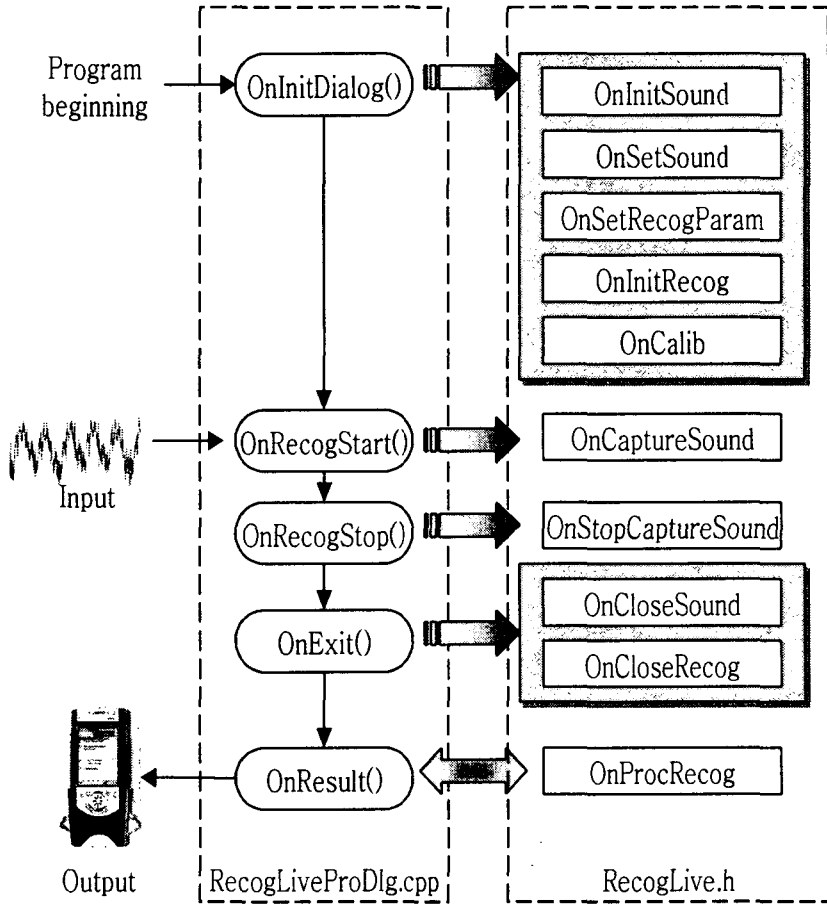
PC	Pentium 150MHz 이상	
OS	<ul style="list-style-type: none"> · Windows 98 second Edition · Windows NT workstation 4.0 (SP5) · Windows 2000 	
RAM	· Windows 98	24 MB 이상
	· Windows NT · Windows 2000	32 MB 이상
Hard-disk space	Minimum Installation	360MB
	Complite Installation	720 MB

본 논문에서 포켓 PC의 응용프로그램을 개발하기 위해 embedded Visual C++을 사용하였다. 그림 4에서처럼 X86 Debug 모드로 프로그램을 작성하여 데스크탑 PC에서 먼저 에뮬레이터를 실행한 후 프로그램 이상 유무를 체크한다. 개발한 프로그램이 에뮬레이터에서 가동되면 ARM Release로 다시 프로그램을 작성하고 USB Sink 케이블을 통하여 PDA의 시작메뉴에 추가하게 되어 PDA상에서 프로그램이 작동된다.



<그림 4> PDA에 대한 SDK 블럭도

4.2. 응용프로그램 알고리즘



<그림 5> 응용프로그램 알고리즘

그림 5는 응용프로그램의 알고리즘을 나타낸 것이며 처리과정은 다음과 같다. 먼저 PDA 상에서 프로그램을 실행하면 WINAPI에 의해 음성인식을 위한 파라미터와 메모리 등의 초기화가 이루어지는 `OnInitDialog()` 함수가 호출된다. 인식 버튼을 클릭하면 마이크를 통해 녹음된 음성을 캡처해 파라미터 값을 음성인식 루틴 함수로 보내지고 여기서 음성 DB와 비교하여 결과를 PDA 화면에 음성 파형과 인식된 결과 값을 출력한다. 마지막으로 인식 중지 버튼을 클릭하면, 음성캡처, 사운드 음성인식을 위한 메모리가 제거된다 [9].

그림 6은 응용프로그램 소스 일부분이며 여기에 사용되어진 각 함수의 의미는 표 4와 같다.

```

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(TRUE);
    return TRUE;
}

```

<그림 6> 응용프로그램 소스 일부

<표 4> 함수의 의미

함수	의미
OnInitSound	Sound 초기화 함수
OnSetSoundParam	파형그리기 함수
OnsetRecogParam	음성인식을 위한 파라미터 초기화
OnInitRecog	음성인식을 위한 메모리초기화
OnCaptureSound	음성캡처 시작함수
OnProcRecog	음성인식 루틴 함수
OnStopCaptureSound	음성캡처 종료함수
OnCloseSound	Sound 종료함수
OnCloseRecog	음성인식을 위한 메모리 제거

5. 실험결과 및 고찰

5.1. 실험환경

본 실험을 위해 PDA는 컴팩사의 모델을 사용하였고, 빠른 데이터 전송과 음성 처리를 위해 기본 메모리에 CF 메모리를 추가하여 실험하였다. 또한 PDA에 대한 응용프로그램의 개발을 위해 데스크탑과 PDA사이에 ActiveSync 3.1로서 서로 동기

화 하였으며, 표 5는 PDA와 데스크탑에 대한 실험환경을 나타낸 것이다.

<표 5> PDA와 데스크탑 PC에 사용된 H/W 사양

* PDA	
Model	iPAQ H3660
CPU	Intel 32bit strongARM 206MHz
Memory	Main(64M) + CF Memory(128M)
Platform	Windows CE 3.0
SDK	embedded Visual C++
* Desktop	
Processor	Intel Pentium III 733MHz
MainMemory	256MB
O/S	Windows 2000

5.2. 인식실험 및 고찰

5.2.1. 인식실험

본 연구에 사용된 가변어휘 음성인식을 위해 사용한 음성 DB는 서울과 서울근교 지명, 도로명, 상호 그리고 증권명 등 총 1074 단어를 인식대상 단어로 사용되었다. 20대 성인 남성화자 70명을 대상으로 52명은 학습, 18명은 인식실험을 위한 데이터로 사용하였다.

먼저 인식실험을 위해 자동차 항법장치에서 자주 사용되는 22개 단어를 선정하여 Stand-alone 모델에서의 가변어휘 인식실험을 TEST1, PDA 상에서의 인식실험을 TEST2로 하여 각각 22(단어/명)×18명=396개 단어를 인식실험 하였다. 또한 PDA에 대한 인식실험을 위해 PDA의 사운드 특성상 장착된 마이크에서 약 10cm 떨어진 위치에서 조용한 실험실 환경에서 인식실험을 하였다. 표 6는 인식실험에 사용한 단어들의 리스트를 나타낸 것이다.

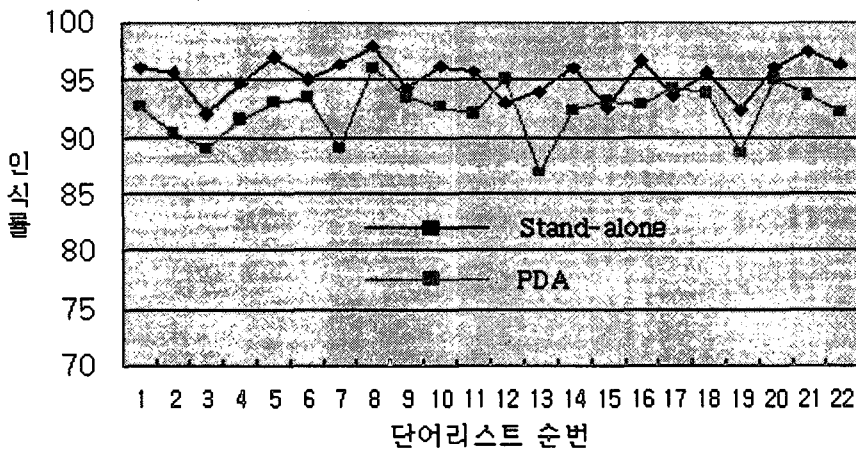
<표 6> 인식실험에 사용한 단어 리스트

순번	1	2	3	4	5	6	7	8	9	10	11
단어	메뉴명	뉴스	메인 메뉴	정치	경제	사회	스포츠	방송 정보	표준 FM	음악 FM	연예 정보
순번	12	13	14	15	16	17	18	19	20	21	22
단어	문화 정보	증권 정보	종합 지수	종목 시세	등락 종목	투자 정보	교통 정보	교통 하나	교통 들	교통 셋	교통 넷

표 7은 각 모델에 대한 인식률을 나타낸 것이다. 그림 7은 각각의 단어 리스트에 대한 인식률을 나타낸 것이다.

<표 7> 가변어휘 인식기반에서의 각 모델에 대한 인식률

Model	인식수/실험횟수	인식시간(초)	인식률(%)	비고
Stand-alone (TEST1)	376/396	단어당 평균 0.8초	95%	DB용량은 서로 같음
PDA (TEST2)	364/396	단어당 평균 2초	92%	



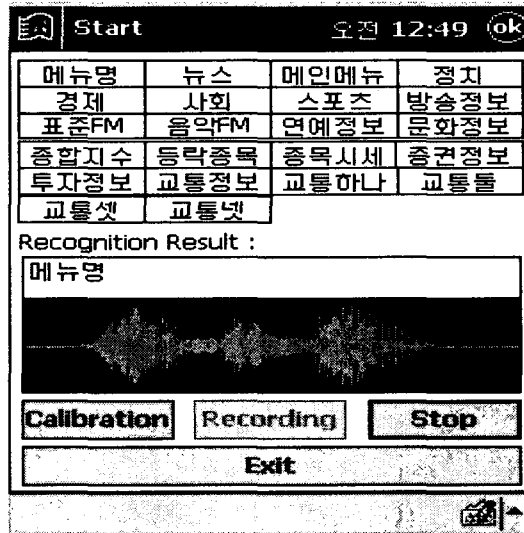
<그림 7> 단어리스트 순번에 대한 인식률

다음 표 8는 Stand-alone 모델에서의 Mixture 수 변화에 따른 인식실험 결과를 나타낸 것이며 4개의 Mixture일 경우가 3개일 경우보다 약 1%의 성능향상을 보였으나 인식시간이 1.5배 증가하였다.

<표 8> Stand-alone 모델에서의 Mixture 수에 따른 인식실험

Mixture	인식수/실험횟수	인식률	실험환경	마이크로폰
3	376/396	95%	컴퓨터실	다이내믹
4	380/396	96%	컴퓨터실	다이내믹

그림 8은 “메뉴명” 단어에 대한 PDA 화면을 나타낸 것이다.



<그림 8> 메뉴명에 대한 PDA 화면

5.2.2. 고찰

본 논문에서는 PDA 상에서 음성 제어를 구현하기 위해 가변어휘 음성인식기를 설계하였고 PDA의 SDK로는 embedded Visual C++을 이용하여 응용프로그램을 개발하였다. 또한 데스크탑에서의 Stand-alone 모델과 비교하여 인식실험을 하였다. 실험결과에서 Stand-alone 모델의 가변어휘 음성 인식률은 PDA상에서의 인식률보다 약 3%, 인식속도 면에서 약 2.5배 정도의 성능 향상을 가져왔다. 그리고 Stand-alone의 경우 mixture가 3인 경우가 4인 경우보다 인식률은 떨어지지만 속도 면에서는 더 우수하다는 것을 알 수 있었다. 하지만 PDA 상에서의 인식률이 비록 Stand-alone 모델보다 떨어지더라도 가변어휘 음성인식기를 이용하여 음성을 제어함으로써 PDA의 주목적인 휴대성과 이동성을 극대화하였다.

6. 맺음말

본 논문은 PDA의 이동성과 휴대성을 극대화하기 위해 트라이폰 기반의 가변어휘 음성인식기를 설계하여 음성제어를 구현하였다. 응용프로그램의 개발을 위해

Embedded Visual Tools 3.0의 embedded Visual C++을 사용하였다.

인식실험을 위해 가변어휘 인식기는 결정트리 상태공유 알고리즘을 기반으로 설계되었으며, Stand-alone 모델과 PDA에 적용시켜 비교실험을 하였다. 이를 위해 20대 성인 남성화자 70명을 대상으로 52명은 학습, 18명은 인식실험을 하였다. 그 결과 Stand-alone 모델이 PDA보다 인식률과 인식속도 면에서 모두 우수한 결과를 나타내었다. 그러나 프로그램을 갱신과 이식할 때 문제점을 갖고있는 Stand-alone 모델에 비해 PDA는 언제 어디서나 사용자가 원하는 최신 프로그램을 제공 할 수 있는 장점을 가지고 있다.

PDA상에서의 음성인식기 구현은 3가지의 문제점을 갖고있다. 첫째, 음성입력에 대한 응답이 데스크탑보다 입력이 지연되거나 손실되는 현상이 있고, 둘째, PDA 자체의 AGC 기능 때문에 잡음환경에 대한 적응이 어렵고, 셋째, 응용프로그램 개발시 에뮬레이터 상에서의 결과와 실제 PDA상에서 실행한 결과가 일치하지 않는 경우가 있다.

향후 이러한 문제점들을 보완하여 PDA상에서의 인식률과 인식속도를 Stand-alone 모델 수준으로 향상시켜 PDA를 클라이언트로 데스크탑을 서버로 하는 모바일 컴퓨팅에 적합한 C/S 모델을 적용시켜 실험을 할 계획이다.

참 고 문 헌

- [1] 배찬권(2001), 세계 PDA 시장 전망, KISDI IT FOCUS, pp.61~65.
- [2] <http://www.microsoft.com/korea/msdn/msdnmag/issues/01/01/cetools/>
- [3] 이재왕(1999), 인터넷상에서의 한국어 음성인식, 석사학위논문, 동신대학교.
- [4] 서봉수(2001), 가변 어휘 음성 인식기 구현 및 탐색 시간 단축 알고리즘 비교, 석사학위 논문, 전남대학교 전자공학과.
- [5] Steve Young (1999), *The HTKBook (for version 2.2)*, Entropic LTD.
- [6] 김동화(1999), 연속음성인식을 위한 향상된 결정트리 기반 상태 공유 기법 연구, 박사학위 논문, 부산대학교 전자계산학과.
- [7] <http://www.microsoft.com/korea/msdn/library/techart/>
- [8] <http://msdn.microsoft.com/vstudio/device/>
- [9] 박상훈, 김철, 최승호(2001), PDA상에서 음성명령어를 구현하기 위한 음성인식기의 설계, 「한국음향학회추계학술발표대회」, 한국음향학회, pp.37~40.

접수일자: 2002년 5월 7일

게재결정: 2002년 5월 24일

▶ **곽상훈(Sang Hun Kwak)**

주소: 전남 나주시 대호동 252

소속: 동신대학교 대학원 정보통신공학과 석사과정

전화: 061-330-3775, 011-9642-8454

E-mail: stepara@empal.com

▶ **최승호(Seung Ho Choi)**

주소: 전남 나주시 대호동 252

소속: 동신대학교 멀티미디어통신공학전공

전화: 061-330-3194

E-mail: shchoi@white.dsu.ac.kr

▶ **신도성(Do Sung Shin)**

주소: 광주시 북구 용봉동 300

소속: 전남대학교 대학원 전자공학과 박사과정

전화: 062-530-0472, 016-608-3736

E-mail: bomb001@hanmail.net

▶ **김진영(Jin Young Kim)**

주소: 광주시 북구 용봉동 300

소속: 전남대학교 전자공학과

전화: 062-530-1757

E-mail: kimjin@dsp.chonnam.ac.kr