PDA상에서 음성인식을 이용한 차량번호 조회시스템

김 우 성, 김 동 환, 윤 재 선, 홍 광 석
성균관대학교 전기전자 및 컴퓨터공학부 휴먼컴퓨터 연구실

A car number retrieving system using speech recognition for PDA

Woo-Sung Kim, Dong-Hwan Kim, Jeh-Seon Youn, Kwang-Seok Hong
HCI Lab, Electrical & Cumputer Engineering, Sungkyunkwan University

maxcel@hanmail.net, kdh2k@popsmail.com, sunhci@ece.skku.ac.kr, kshong@yurim.skku.ac.kr

요약

본 논문에서는 PDA상에서 음성인식과 합성을 통하여 차량 번호를 조회할 수 있는 시스템을 구현하였다. 차량번호 인식을 위한 4연속 숫자음과 명령어 인식부분, 그리고 각 단계별로 합성된 음성을 들려주도록 구성하였다. 본 연구의 인식시스템은 화자독립으로 실험을 하였으며, 여러화자에 대한 4연속 차량번호 인식률과 명령어에 대한 인식률은 각각 97%, 99%가 나왔다.

Abstract

In this paper, we present a car number retrieving system using speech recogntion and speech synthesis for PDA. This system consist of 4-digit numbers and command speech recognition as well as speech synthesis. Experiment results showed 4-digit numbers recognition rate 97% and commands recognition rate 99% through speaker-independent method.

I. 서론

PDA상에서 음성인식을 이용한 차량번호 조회시스템은 요즘에 PDA시장의 확대와 더불어 음성인식 기술을 접목하여 보다 입력을 쉽게 처리하기 위한 기술이다. 작은 PDA상에 문자나 숫자를 터치스크린에 눌러 쓰는 것보다 직접적으로 음성으로 입력하면 쉽게 처리할 수 있기 때문에 이것을 가지고 자동차 차량번호를 조회하는 프로그램을 구현하였다.

PDA의 OS는 Window CE를 사용하였으며, PDA상에 달려있는 마이크와 함께 외부 헤드셋을 통한 마이크로 입력을 받았다. 본 연구는 PDA에서 음성인식과 합성을 통한 인간과 컴퓨터간의 보다 편리한 인터페이스를 위하여 음성인식을 이용한 차량번호 조회 시스템을 구현하고 그 성능을 확인하였다.

II. 차량번호 조회시스템의 구성

본 차량번호 조회시스템에서 인식 시스템 구성에 필요한 명령어는 크게 숫자와 선택 명령어가 있는데 숫자는 "공, 일, 이, 삼, 사, 오, 육, 칠, 팔, 구"와 선택 명령어로는 "입력/수정, 검색/수정, 예/아니오, 선택숫자(리스트 중에서 선택하는 숫자)/검색, 검거/취소, 확인/취소"가 있다. 음성 입력이 들어오면 이중에서 확률적으로 높은 것이 인식되며 그리고 각 단계별로 합성음성을 통하여 안내 메시지가 나온다. 그리고 PDA로 받은 숫자 데이터를 보충하기 위해서 Desktop PC에서 받은 숫자 데이타를 추가하였다.

숫자음성 데이터 베이스는 표 2.1과 같다.

표 2.1 숫자음 데이터 베이스

PDA로 녹음(숫자,메뉴)		Desktop PC로 녹음(숫자)	
데이타명	화자수	데이타명	화자수
PDA1	70명	PC	230명
PDA2	29명		
PDA3	27명		
PDA4	50명		

차량조회 시스템은 그림 2.1과 같은 과정을 거치게된다.

(1) 프로그램의 초기화면에 근무자 ID와 비밀번호에 각각 4연속숫자를 발음한다. 만일 제대로 입력이 되었다면 "입력"을 발음을 하고 틀렸을 경우 "수정"을 발음한다. 이 단계에서 합성음성이 나오는데 "사용자 확인입니다. 근무자 아이디를 말씀하십시오. 비밀번호를 말씀하십시오"와 같이 나온다. 또한 아래 각 단계별로 안내하는 합성음성이 나온다.
(2) 수배차량 조회단계에서는 차량번호 또는 메뉴 바꿈을 선택할 수 있는데 메뉴 바꿈시에는 당일검거 차량조회단계로 넘어간다. 차량번호 선택시 검색/수정을 선택을 한다.
(3) 입력값 확인 단계에서는 예/아니오 선택한다.
(4) 검색결과 단계로써 같은 자동차 번호목록이 나온다. 사용자가 여러 개의 동일 자동차 번호 중에서 임의의 리스트 번호에 대해서 조사할 수 있고 또는 다시 수배차량조회 단계로 넘어갈 수 있다.
(5) 자세한 사항단계에서는 검색 리스트 중에서 선택된 자동차 번호에 관한 자세한 사항들이 표시된다. "검거" 선택시에는 다시 수배차량 조회단계로 "취소"시에는 검색결과 단계로 넘어간다.

III. 음성인식과 음성합성

본 차량조회 프로그램은 syllable단위 또는 word단위로 음성인식을 하였다. 합성음은 미리 정하여진 안내 메시지를 상용화 프로그램을 이용하여 만든 합성음을 가지고 편집하여서 각 단계별로 적절한 위치에 합성음이 나오도록 하였다.

3.1 차량번호 조회에서의 음성인식

본 논문에서는 제한적인 숫자음(공, 일, …, 구)과 몇개의 명령어 단어를 가지고 음성인식을 한다.

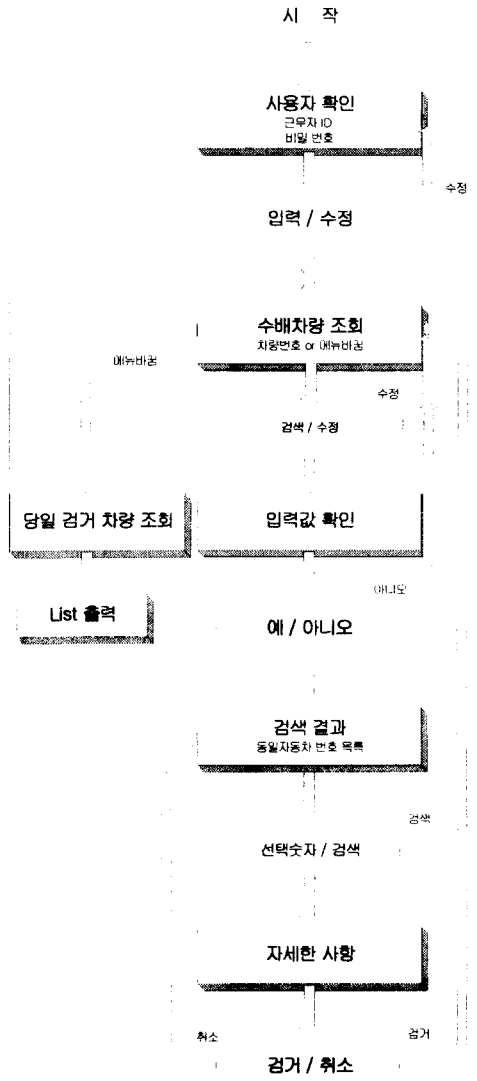

그림 2.1 차량조회 프로그램의 순서도

숫자음 인식을 위해서 총 406명의 화자, 그리고 명령어에 대한 인식을 위해서 176명의 화자로부터 수집한 음성으로 reference model을 구성하였다. 인식은 음성입력이 들어오면 Mel-Cepstrum 파라미터 추출을 거치고 Vector Quantizer를 통과하여 128 Level중 하나의 값을 가지는 sequence형태로 표현되며 DHMM을 거쳐서 인식을 하게 된다. 여기서 DHMM 모델중 확률적으로 큰 값을 가지는 것을 인식 결과로 나오게 된다.
이 과정은 그림 3.1에 나타내었다.

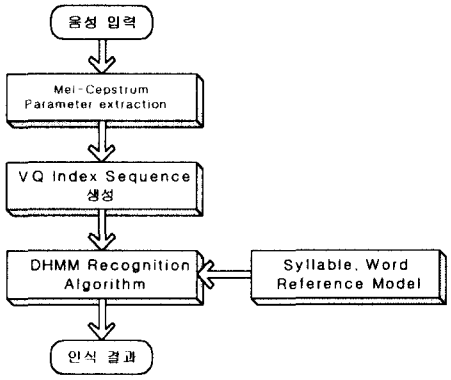

그림 3.1 차량번호 조회 프로그램에서
음성인식 처리과정 [1]

차량번호 조회 프로그램에서는 4연속숫자에 대한 인식을 수행한다. 각 숫자음절에 대한 평균길이가 다르므로 짧게 발음되는 음절에는 Duration(state수)을 짧게 할당을 해주고 그리고 길게 발음되는 곳에는 비교적 작은 상태를 할당해주어서 전체적으로 인식률을 높여주는 과정을 사용하였다.

그림 2.1에 수배차량 조회단계에서 수배차량 번호 또는 메뉴 바꿈이라는 두 개중에 하나의 음성인식 판별과정이 필요한데 두 명령음성의 음절수가 4음절이라서 보다 정확한 인식을 위해 입력음성이 어떤 4연속숫자 또는 "메뉴 바꿈"이라는 음성이 들어오면 숫자인식 model part와 메뉴인식 model part 두 부분에 모두 들어가 나온 인식확률 중 임계치를 넘는 쪽으로 인식결과를 만든다.

3.2 차량번호 조회를 위한 음성합성

본 프로그램에서 음성 인식된 상태를 알려주고 사용자에게 확인을 위해서 음성합성을 이용하였다. 여기서 사용된 합성음은 각 단계에서 필요한 음성을 상용화 프로그램을 이용해 몇 개의 문장을 합성해서 단계별로 필요할 때마다 합성된 음성을 들려주는 방법을 쓰고 있다. 합성기를 통한 음성보다는 직접 녹음한 음성을 사용하면 보다 자연스러운 음성을 들을 수 있으나 여기에서는 합성음을 사용하였다.

Ⅳ. 실험 및 결과

본 실험에서는 5명의 화자가 4연속 숫자 15개를 발음하였을 경우와 명령어 선택시 음성인식 결과를 실험하였다. Database는 PDA에서 음성을 받은 것과 Desktop PC에서 받은 숫자음을 가지고 작업을 하였다. 표2.1에서 데이터명 PDA1~PDA3까지는 PDA에 내장되어 마이크를 통하여 숫자와 명령어 음성을 받았고 PDA4는 따로 제작한 헤드셋 마이크를 가지고 녹음을 하였다. PDA녹음은 실내와 실외에서 녹음을 하였고 Desktop PC 녹음은 실험실에서 Dynamic 마이크를 가지고 녹음하였다. 녹음시 11.025kHz로 샘플링을 하였다.

일반적으로 LPC(Linear Predictive Coding)보다는 LPC-Cepstrum이 보다 좋은 인식 성능을 가지고 있으며 이것보다 잡음에 더 강한 특성을 가지고 있는 Mel-Cepstrum을 이용하면 인식률이 좋다[4]. 본 실험에서 사용된 4연속숫자 15개의 set을 보면{1427, 0287, 3510, 4621, 8194, 7954, 8065, 9176, 9205, 9601, 1199, 6633, 2244, 5500, 8877}으로 /공/~/구/까지 균등하게 분포시켰으며 한 화자가 4개숫자×15개 = 60개 발음한 것을 가지고 인식률을 계산하였다. 그리고 명령어에 대한 인식결과는 각 단계별로 두개 명령어 중에 선택하는 것이다. 실험한 결과를 보면 MFCC를 사용한 경우 4연속숫자 차량번호 평균 인식률이 97%이다. 표 4.1에 인식 결과를 나타내었다.

표 4.1 Mel-Cepstrum을 사용한
4연속 차량번호 인식결과

화자	인식된 숫자 개수															총수	인식률(%)
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
화자1	1	1	1	4	1	1	1	4	1	1	4	1	4	4	1	60	100
화자2	1	3	1	4	3	4	4	4	1	1	3	4	4	4	4	58	96.7
화자3	1	3	4	1	4	4	4	4	4	4	4	3	4	4	4	58	96.7
화자4	3	1	1	1	1	1	4	4	4	4	3	4	4	4	1	59	98.3
화자5	1	3	1	4	3	1	1	4	3	3	4	4	4	3	4	56	93.3

인식시 PDA(CPU:200MHz)에서 걸리는 시간을 측정하여 보면 6초정도가 소요되었다. LPC-Cepstrum을 사용할 때 3초정도 소요되는 것에 비해 좀 더 시간이 걸렸다. MFCC가 LPC-Cepstrum보다 계산량이 많아 시간적으로 오래걸림을 알 수 있다.

표 4.2는 그림 2.1에 나타난 각 단계별에 있는 메뉴 명령어에 대한 인식결과를 보여주는 표이다. 실험은 단계별로 4번씩하였는데, 예로써 단계1에서 입력/수정을 선택하는데 "입력"을 두 번, "수정"을 두 번 말할 때 정확히 인식되는 수를 넣었다. 표4.2에서 각 단계에서 선택하는 메뉴가 두가지 밖에 없어서 높은 인식률을 가지는 것을 알 수 있다.

표 4.2 Mel-Cepstrum을 이용한 메뉴명령어 인식결과

화 자	단계1	단계2	단계3	단계4	단계5
	입력/수정	검색/수정	예/아니오	숫자/검색	검색/취소
화자1	4	4	4	4	4
화자2	4	4	4	3	4
화자3	4	4	4	4	4
화자4	4	4	4	4	4
화자5	4	4	4	4	4
인식률(%)	100	100	100	95	100

단지 단계4에서 동일 차량번호 리스트 번호를 선택함에 있어서 약간의 인식오류가 있기는 했다. 연속된 숫자음의 발성시에는 숫자와 숫자 사이에 파형이 연결되어 나타나게 된다. 연속 음성시 단어 단위로 인식을 한다면 그 표준패턴을 위한 기억 용량 및 계산시간이 많아지게 된다[2]. 그러나 본 논문에서는 각 숫자음을 발음함에 있어서 연결 숫자음을 발음하고, 한 개의 음절 단위로 인식결과를 나타내었다.

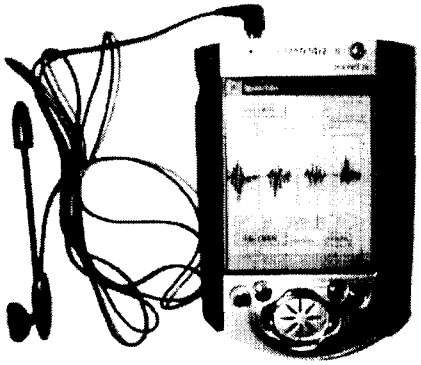

그림 4.1 4연속 숫자인식 프로그램 PDA실행화면

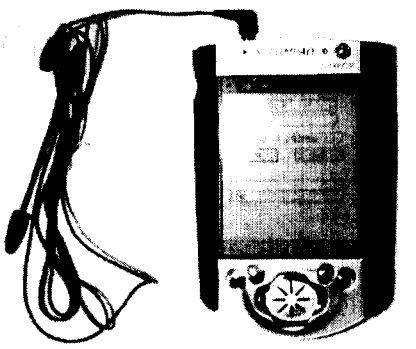

그림 4.2 차량번호 조회프로그램 PDA실행화면

그림 4.1과 4.2에는 구현한 PDA상의 연결숫자 음성인식과 차량번호 조회시스템의 실행화면을 나타내었다.

V. 결론

요즘 PDA의 시장이 급속히 확대되고 있다. 본 논문은 PDA입출력을 위해 음성인식을 접목시켜서 보다 인간에 편리한 인터페이스를 제공할 수 있는 차량번호 조회 프로그램을 구현하여 보았다.

4연속숫자에 대한 인식률은 97%정도로 높게 나왔으며 메뉴 명령어에 대한 전체적인 인식률은 99%정도로 높게 나왔다. 메뉴명령어 인식률이 높게 나오는 이유는 그림 2.1에 있는 각 단계별로 두 개의 명령어에 대한 인식이므로 조금 높게 나왔다. 또한 프로그램 실행하는 환경을 비교적 조용한 환경에서 수행하였기 때문에 인식률이 높게 나왔다.

하드웨어적인 면, 즉 외부 헤드셋 마이크 성능, PDA내부 하드웨어도 이번 실험에서 중요한 면으로 작용한다는 것을 알 수 있었다.

잡음처리 문제, 숫자연음 발음시 처리문제, 음절길이 조절문제등이 연구과제로 남아있다. 실제적으로 PDA가 실외에서 많이 사용된다는 점을 감안한다면 잡음처리 문제가 가장 큰 걸림돌으로 작용한다. 이런 문제들이 해결되었을 때 PDA를 통한 음성인식 정확도가 높아질 수 있고 PDA에 음성인식 및 합성이 기본 인터페이스로 자리매김을 할 수 있을 것이다.

* 본 연구는 (주)스피치이텍의 연구비 지원으로 이루어졌습니다.

참고문헌

[1] 김동환, 윤재선, 홍광석, '음성인식 끝말 이어가기 게임의 구현,' 한국 신호처리·시스템학회 논문집 제1권 제2권, pp177-180, 2000.

[2] 윤재선, 'Unrestricted continuous digits speech recognition using semisyllable HMM juncture,' 석사학위 논문, 성균관대학교 전자공학과, pp16-30, 1997.

[3] 윤재선, 홍광석, '반음절 단위 HMM을 이용한 연속숫자 음성인식,' 한국음향 학회지 제17권 제5호, pp73~78, 1998.

[4] Douglas O'Shaughnessy, 'Speech Communications-Human and Machine,' IEEE Press, 2000.

[5] Claudio Becchetti and Lucio Prina Ricotti, 'Speech Recognition Theory and C++ Implementation,' John Wiley & Sons, 1999.