제15회 음성통신 및 신호처리 워크샵(KSCSP '98 15권1호)

웹 브라우저 상에서 한국어 음성인식을 이용한 정보검색 시스템

이항섭, 김회린

한국전자통신연구원 음성신호처리팀

Information Retrieval System Using Korean Speech Recognition on the Web Browser

Hang-seop Lee, Hoi-rin Kim

Speech Signal Processing Technical Team, ETRI
hslee@etri.re.kr, hrkim@zenith.etri.re.kr

요약

본 논문은 웹 브라우저 상에서 한국어 음성인식을 이용한 정보검색 시스템에 대한 것이다. 이 시스템의 특징은 웹 브라우저 상에서 보여지는 HyperText Word를 인식할 수 있는 것으로 기존의 웹 브라우저를 마우스 click 대신 음성인식을 이용하여 사용할 수 있다는 것이다. 웹 브라우저를 통해서 보여지는 고정되지 않고 계속하여 변화하는 인식후보를 인식하기 위해 당 연구실에서 개발한 가변 어휘 인식기를 사용하였다. 시스템은 Windows 95/NT 환경에서 개발되었으며, 사용자가 새로운 인터페이스를 배울 필요 없이 바로 사용할 수 있도록 사용자 편의성 부분도 고려하여 개발되었다. 개발된 시스템은 독립 환경, 독립 화자에 대해 실험한 결과 130여개의 단어에 대해 평균 90% 정도의 인식성능을 보인다.

1. 서론

인터넷의 소개 이후 인터넷의 사용은 지금도 전세계적으로 급격히 늘어나고 있으며 1998년 현재 전세계의 인터넷 호스트는 2,967만대에 이르며, 그 사용자는 3억에 달하고 있다. 인터넷은 정보 검색 및 정보 공유의 표준적인 방법으로 사용되고 있다. 국내에서도 인터넷 이용자는 급격히 증가하고 있으며, 그 영역은 전문적인 내용에서부터 기업 소개, 홈 쇼핑, 레저, 문화 및 전자 상거래로 계속 확장되고 있는 추세이다. 인터넷 뿐만 아니라 인트라넷의 사용도 사내 전자결재 시스템의 도입으로 크게 확산되고 있는 실정이다. 이러한 인터넷과 인트라넷의 사용은 웹 브라우저를 통해 이루어진다. 이러한 웹 브라우저에 음성을 이용한 인터페이스가 추가 된다면 보다 쉬운 인터넷의 사용으로 사용자 층을 더욱 넓힐 수 있을 것이다.

이러한 이유에서 우리는 음성인식을 이용하여 웹 브라우저 상에서 정보를 검색할 수 있는 시스템인 EVANS(ETRI Voice Activated Navigation System)를 개발하였다. 음성인식을 사용하여 웹 브라우저 상에서 정보검색을 하는 데는 아직 여러 가지 문제점들이 남아 있지만 우리의 이러한 시도가 음성인식의 실용화를 앞당길 수 있기를 기대한다.

Section 2 에서는 EVANS 에서 사용한 가변어휘 인식기에 대해 간단한 소개를 하고, section 3 에서는 EVANS 의 개념과 특징을 소개한다. section 4 에서는 EVANS 시스템의 구성과 각 모듈 및 그 구현 방법을 설명하며, section 5 에서는 on-line 상에서의 인식실험 결과를 소개한다.

2. 가변어휘 인식기

가변어휘 인식기[1][2]의 구조는 그림 1 과 같다. 이 인식기는 미리 훈련되지 않은 단어를 인식할 수 있는 기능을 가지고 있으며, 인식후보 단어가 바뀌게 되면 새로운 단어에 대한 훈련 없이 발음사전만 바꾸는 것으로 새로운 인식후보를 인식할 수 있다. 가변어휘 인식기는 POW 3,848 단어 데이터[3]를 사용하여 훈련되었다.

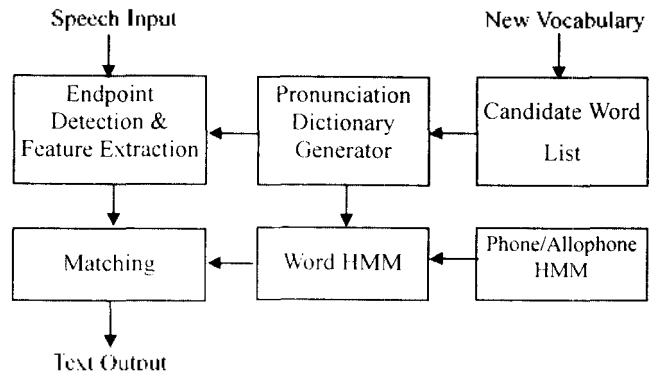

그림 1. 가변어휘 인식기의 구조

3. EVANS

EVANS는 웹 브라우저 상에서 음성인식을 이용한 정보검색을 제공하는 시스템으로 EVANS의 특징은 다음과 같다.

1) Browser 메뉴에 대한 음성 명령 :
사용자는 브라우저 윈도우 control을 포함한 모든 웹 브라우저의 메뉴를 음성으로 접근할 수 있다. 예) "인쇄", "열기", "최대크기", "스크롤업"
2) Link 에 대한 음성 명령:
사용자는 현재의 웹 페이지가 제공하는 모든 Link 들을 음성으로 접근할 수 있다.
3) 음성북마크
사용자는 음성으로 접근할 수 있는 자신만의 음성 북마크를 만들 수 있다.
4) 음성명령을 현재의 Web page 에 표시 :
사용자는 인식할 수 있는 음성인식 후보를 웹 페이지 상에서 볼 수 있다.
5) 메뉴와 음성북마크에 대한 음성명령 재 정의 :
사용자는 웹 브라우저 메뉴와 음성북마크에 대한 음성명령을 사용자가 원하는 대로 재 정의 할 수 있다.
6) 화자독립, 어휘독립 인식

EVANS는 두 가지 기능을 가지고 있다. 하나는 웹 브라우저가 현재 보여주고 있는 웹 페이지에서 인식후보를 추출하여 새로운 단어 사전을 생성하는 것이고, 또 다른 하나는, 인식결과에 따라 웹 브라우저를 동작시키는 것이다. EVANS가 동작되면 음성에 의해서 웹 브라우저를 동작시키기 위한 음성인식후보를 얻기 위해 현재의 웹 페이지에 대한 parsing을 수행한다. 웹 페이지에 대한 parsing 결과로 얻어진 인식후보은 HTML 파일에 가록되고 웹 브라우저를 통해 보여진다. 이제 사용자는 자신이 가고자 하는 site로 가지 위해서 그 site를 나타내는 하나 또는 몇 개의 단어를 발성하면 된다.

다른 곳으로의 link를 제공하는 HyperText word 들은 웹 페이지에 따라 계속하여 변화하기 때문에 이들을 인식하기 위해서 EVANS에서는 미리 훈련되지 않은 단어들에 대한 인식이 가능한 가변어휘 인식기를 사용하였다.

4. 시스템 구성 및 구현

4.1 시스템 구성

EVANS는 그림 2와 같이 HTML parser, 인식후보 추출 모듈, 가변어휘 인식기, 인식결과 분석기로 이루어졌다. 웹 브라우저가 새로운 HTML 파일을 open 하면 이 파일은 HTML parser로 전달된다. HTML 파일은 다른 site로의 접속을 제공하기나 같은 페이지내의 특별한 곳으로 이동하기위한 link(URL)를 가지고 있다 인식후보를 추출하기 위해서 이러한 link 들을 가지는 hypertext word를 사용하였다. Hypertext word들은 서로 다른 길이를 가지며, 단어, 이미지, 숫자 또는 이런 것들의 혼합으로 이루어진다. 이러한 다양한 형태의 hypertext word 중 부터 실제로 사용될 인식후보를 추출하기 위해 "인식후보 추출 모듈"이 사용되었다. 인식후보 추출 모듈은 hypertext word들과 음성북마크 명령, 웹 브라우저 메뉴, 웹 브라우저 윈도우 control 명령을 사용하여 현재의 웹 브라우저 상에서 인식 가능한 인식후보를 만들어 낸다.

음성북마크는 웹 브라우저가 제공하는 북마크와 같은 기능을 제공하는 것으로 사용자가 음성으로 미리 등록된 site에 접속할 수 있게 해준다. 또한, 사용자는 음성북마크 파일의 내용을 text 편집기를 사용하여 음성으로 접속하고자 하는 site를 쉽게 추가, 삭제 및 수정할 수 있다.

표 1. 웹 브라우저 윈도우 control 명령

Browser Window Control Command	
Normalize	Scroll up
Maximize	Scroll down
Minimize	Scroll right
Page up	Scroll left
Page down	

HTML parser로 부터 얻어지는 hypertext word로 부터 음성인식 후보를 생성하는 과정은 다음과 같다. 첫째, 만일 hypertext word가 3 단어 이상으로 이루어져 있으면 3 단어까지만 추출하며 추출된 단어들을 모두 붙여서 하나의 단어로 취급을 한다. 둘째, 이미 똑같은 이름의 인식후보가 존재 한다면 이를 "하나", "둘", "셋" 과 같은 연속적인 숫자 인덱스로 바꾸어 준다. 또한 이러한 숫자 인덱스는 그림이나 숫자 또는 한국어를 제외한 다른 언어 등 hyperlink를 제공하는 것이 한글이 아닌 경우에 대해서도 적용된다. 이 과정을 거쳐 생성된 인식후보들은 각 인식후보들이 가르키는 link(URL)과 함께 인식후보 리스트에 기록된다. 가변어휘 인식기는 현재의 웹 페이지에서 추출된 인식후보 리스트를 사용하여 인식을 수행하며 인식된 결과는 인식결과 분석기로 전달된다. 인식결과 분석기는 인식기의 인식결과에 해당하는 적절한 명령을 웹 브라우저에 보낸다. 웹 브라우저로는 Netscape Navigator를 사용하였다. Netscape Navigator와 통신을 위해서는 Dynamic Data Exchange (DDE) method [4]를 사용하였다.

4.2 시스템 구현

EVANS는 Windows 95/NT 환경에서 Pentium pro 200MHz CPU, 일반적인 사운드 카드, Netscape Navigator를 갖춘 PC에서 구현되었다. EVANS는 자동음성검출 [5] 기능을 가지고 있어 음성 입력의 시작과 끝을 시스템에 알리기 위한 keyboard나 mouse 입력을 필요로 하지 않는다. 이 기능은 사용자에게 편의성을 제공하는 반면, 주변 소음이 너무 크거나 일정하지 않은 상황에서는 시스템의 성능에 나쁜 영향을 주기도 한다. EVANS는 또한 음성을 인터페이스로 사용하기 위한 여러 가지 유구사항[6]들을 갖추고 있다. EVANS에서 사용된 가변어휘 인식기는 ETRI에서 개발한 화자독립, 어휘독립 SCHMM-base의 인식기이다. 이 인식기는 문맥독립 또는 문맥종속의 음향모델을 사용하였으며, POW 3848 데이터 베이스를 사용하여 훈련하였다.

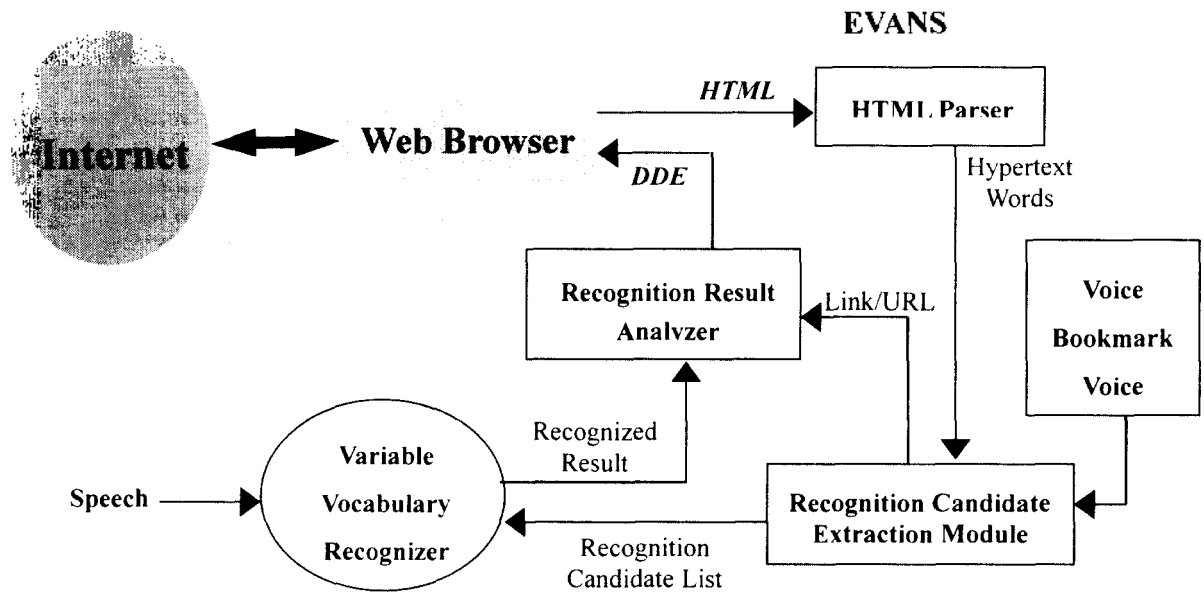

그림 2. EVANS 구성

이 인식가능 문법 없이 간단한 문장을 인식할 수 있으나, 이 시스템에서는 단어인식 기능만을 사용하였다.

5. EVANS 성능 실험

EVANS 인식 성능 테스트는 on-line 가변어휘 인식 상황에서 인식률과 인식 속도의 변화에 중점을 두었다. 실험은 EVANS의 실제적인 사용 상황에서의 인식률 계산을 위해 실제로 존재하는 인터넷 사이트를 접속한 후 인식 대상인 링크와 Netscape 메뉴 그리고, 음성 북마크 명령을 내리고 그 결과를 기록하는 방법으로 실험을 수행하였다.

5.1 시험 절차

인식 실험에 참여한 화자와 시스템 규격 그리고 환경은 다음과 같다.

- ◆ 화자 : 표준말을 사용하는 30 대 초반의 남성 2 인과 30 대 후반의 남성 1 인
- ◆ 시스템 :
  - 화자 1 : Pentium Pro 200MHz Dual CPU, 128Mbyte RAM
  - 화자 2 : Pentium II 266MHz Dual CPU, 256Mbyte RAM
  - 화자 3 : Pentium 166MHz, 32Mbyte RAM
- ◆ 환경 : 다수의 PC 와 Workstation 그리고 printer 가 있는 사무실 환경
- ◆ 마이크 : SENNHEISER HMD 410 (Close-Talk Mic)

실험은 사이트를 바꾸어 가며 총 4 번을 실시 하였다. 각 사이트는 전체 어휘수와 평균 단어의 길이에 변화가 있는 곳으로 선정하여 단어 수와 단어 길이의 변화에 따른 인식률과 인식시간의 변화를 기록하였다.

EVANS 의 인식대상은 크게 3 가지로 Hyperlink 를 제공하는 단어, Netscape 자체 메뉴, 음성북마크 이다. 실험에서는 이 3 가지 인식 대상을 그 수에 비례하게 선택하여 발성하도록 하였다. 실험 결과는 표 2 에 나타내었다.

5.2 실험 결과

실험 결과는 표 2 에 나타내었다.

표 2. On-line 인식 실험 결과

	전체 어휘 수	가변 어휘 수	평균 단어 길이	발성 단어수	인식률 (%)	인식 시간 (sec)
실험 1	129	90	3.76	50	86.7	1.73
실험 2	116	77	4.03	50	88.7	1.70
실험 3	119	80	4.8	50	93.3	1.76
실험 4	146	106	5.11	50	92	1.948
평균	**127.5**	**88.3**	**4.25**	**50**	**90.2**	**1.78**

가변어휘의 수에 비례하여 인식률이 달라지기 보다는 가변어휘의 종류가 훈련된 데이터에 얼마나 유사하는가가 인식률에 더 큰 영향을 주는 것을 알 수 있다. 또한 발성 단어의 길이가 길수록 인식률이 좋은 결과를 보이고 있어 짧은 단어 보다는 긴 단어가 인식시간은 오래 걸려도 인식률은 더 좋음을 알 수 있다.

6. 결론

본 논문에서는 웹 브라우저 상에서의 음성인식을 이용한 정보검색 시스템인 EVANS를 소개하였다. EVANS에서 인식 가능한 대상은 웹 브라우저 자체 메뉴와 브라우저 윈도우 조절 메뉴, 그리고 현재 웹 페이지에서 보여지고 있는 HTML 파일에서 다른 곳으로의 링크를 제공하는 hypertext word 들이다. 인터넷 사이트마다 달라지는 웹 페이지내의 hypertext word 들을 인식하기 위해서 화자독립, 어휘독립 인식기능을 가지는 가변어휘 인식기를 사용하였다. EVANS를 사용함으로써 인터넷을 통한 정보 검색시 mouse와 함께 음성을 이용한 인터페이스가 가능하게 되었다.

가변어휘 인식기의 시뮬레이션 인식률(93.8%)에 비해 on-line 인식률(90.2%)이 나쁘게 나온 이유는 사용환경의 차이와 이에 따른 음성검출기의 결과가 시스템의 전체적인 인식률에 영향을 미치는 것으로 생각된다. 그러므로, 실제로 음성인식 기술이 실생활에 사용되기 위해서는 인식기 자체의 성능도 중요하지만, 인식기의 성능을 저하시키지 않는 전처리 과정에 보다 많은 신경을 써야할 것이며, 이를 위한 연구가 필요하다.

참고문헌

[1] Hoi-Rin Kim and Hang-Seop Lee, "Variable vocabulary word recognizer using phonetic knowledge-based allophone model," *Jour. of Acoustical Society of Korea*, vol. 16, no. 2, pp. 31-35, Feb. 1997.

[2] 김회린, 이황섭, "POW3848 단어 인식기 구현 및 어휘독립 실험," 제 13 회 음성통신 및 신호처리 워크샵 논문집, 13 권 1 호, pp. 127-130, 1996.

[3] Yeonja Lim and Youngjik Lee, "Implementation of the POW (Phonetically Optimized Words) algorithm for speech database," *Proc. of ICASSP*, pp. 89-91, 1995.

[4] "Netscape's DDE implementation," Version 0x00020002, available from the Internet site http://developer.netscape.com/docs/manuals/communicator/DDE/abtdde.htm.

[5] H. S. Lee and M. S. Hahn, "Development of a real-time endpoint detection algorithm," *Proc. of ICSPAT'93*, vol. 2, pp.1547-1553, Sept. 1993.

[6] Alexander I. Rudnicky, "Speech interface guidelines," available from the Internet site http://www.speech.cs.cmu.edu/rspeech-1/air/papers/SpInGuidelines.