論文99-36S-5-12

정보검색 시스템의 음성 인터페이스 구현

(Implementation of the Speech Interface for Information Retrieving System)

金 正 哲 *, 裵 建 星 *

(Jeong Cheol Kim and Keun Sung Bae)

요 약

본 논문에서는 HMM 고립단어인식 기술을 이용하여 정보 사용자들이 윈도즈 환경에서 편리하게 정보를 검색할 수 있는 시스템을 구현하였다. 인식 시스템에서 인식단위로 유사음소모델을 이용하여 인식어의 확장성을 고려하였고 기본모델은 SPHINX 시스템에서 사용하는 형태의 음소모델을 연속분포 HMM으로 구현하였다. 정보검색 도구에서는 기능을 단순화하고 검색절차를 음성으로 출력하도록 하여 사용자의 편의성을 고려하였다.

Abstract

In this paper, we implemented the user interface of the information retrieving system using isolated-word recognition technique. The phoneme-like unit based HMM(Hidden Markov Model) was used considering the expansibility of words to be recognized. Each phoneme-like unit was modeled by a continuous HMM that adopted SPHINX model for state transition. Considering the user-friendliness, the implemented system simplified the searching procedures and informed the user of the next step with speech output.

Ⅰ. 서 론

초기의 컴퓨터는 키보드를 통해서 컴퓨터에게 명령을 하고 모니터 화면을 통해서 출력을 하는 다소 기계적인 형태의 인터페이스가 대부분이었다. 키보드를 다루는 일과 화면에 텍스트로 출력되는 내용들을 살펴보는 일은 초보자가 익숙해지기까지 많은 시간이 걸리는 번거로운 일이었으며, 이러한 결점을 극복해 보고자 진보된 하드웨어를 바탕으로 GUI (Graphic User Interface) 개념이 도입되었다. 이것은 사람에게 보다 익숙하고 인지하기 쉬운 그래픽 형태의 화면을 바탕으로 원하는 부분을 마우스나 키보드의 이동으로 선택할 수 있는 새로운 형태의 인터페이싱 방법으로 초보자라도 금방 이 환경에 익숙해져서 컴퓨터를 사용하게끔 할 수 있는 획기적인 기술로 오늘날의 컴퓨터에는 이미 보편화되기에 이르렀다. 이는 보다 편리한 인터페이스를 원하는 사용자들의 요구를 받아들인 것이라고도 할 수 있으며, 앞으로도 이러한 편리한 인터페이스에 대한 요구는 보다 다양한 형태로 끊임없이 계속되리라 생각된다.

최근에는 음성을 통해 컴퓨터와 인터페이싱하는 방법에 대한 연구가 많이 진행되고 있다. 왜냐하면 음성

* 正會員, 慶北大學校 電子電氣 工學部
(Department of Electronics and Electrical Engineering, Kyungpook National University)

※ 본 연구는 정보통신부 초고속 정보통신 응용 기술 개발사업의 연구비(과제번호 : 96-68) 지원으로 수행되었으며, 지원에 감사드립니다.

接受日字:1998年1月5日, 수정완료일:1999年4月26日

은 인간의 기본적이고 자연스러운 의사 소통 수단이므로 특별한 훈련이 필요없이 기계에 적용하여 보다 편리하고 자연스러운 형태의 정보 입력이 가능하기 때문이다. 최근들어 음성인식기술의 상당한 진보와 이를 수용할 수 있는 컴퓨터 발전으로 인해 음성을 개인용 컴퓨터에 응용할 수 있는 잠재력이 서서히 실현되고 있다[1]. 이미 마이크로소프트사에서도 다음 버전의 운영체제에는 사람의 음성을 듣고 그 명령에 따라 컴퓨터가 동작할 수 있도록 하는 방법, 즉 음성인식기술을 포함할 것이라고 발표한 바가 있다[2].

방대한 양의 정보가 매일 쏟아져 나오는 현대 정보화 사회에서 정보검색 시스템에 음성인식기술을 사용한다면 불특정 다수의 일반 사용자가 정보검색을 편리하게 할 수 있는 인간 중심의 정보검색 서비스를 제공할 수 있을 것이다. 이에 본 연구에서는 제한적인 단어에 대해서 HMM(Hidden Markov Model)을 이용한 화자 독립적인 음성인식 시스템[3-7]을 구현하고, 그에 적합한 정보검색 시스템의 사용자 인터페이스를 구성하였다. 이들 인터페이스는 그래픽 사용자 인터페이스 환경인 윈도즈 운영체제에서 구현되었으며 음성 인터페이스를 위한 음성인식은 HMM을 이용한 고립단어 인식기술[7]을 사용하였다. 인식모델의 기본 단위는 유사음소모델들로 구성되며, 각 기본모델은 SPHINX 시스템[6]에서 사용된 형태의 음소모델을 연속분포 HMM으로 구현하였다. 음소단위 HMM을 사용함으로써 검색어의 확장을 용이하도록 하였으며 다중 화자의 사용이 가능하도록 화자독립적 음성 인터페이스를 구현하였다.

Ⅱ. 음소단위 HMM을 이용한 고립단어 인식

인식 어휘량이 증가하면 혼동하기 쉬운 단어 수가 증가하게 된다. 또 대용량 어휘의 경우 모든 단어에 대응하는 모델을 학습시키기 곤란하므로 subword 단위를 모델링할 필요가 있다. 본 연구에서는 정보검색 시스템의 명령어에 사용되는 모든 음소를 총 41개의 유사음소로 정의하였으며, 이들을 표 1에 나타내었다.

이 유사음소는 일반 음성학에서의 음소와 비슷하지만 몇 가지 음향학적인 효과가 고려되었다. 즉, 자음의 경우 'ㄱ', 'ㄷ', 'ㅂ' 등은 음성의 첫부분에서 발성될 때, 그리고 종성으로 발음될 때를 분리하여 모델링하였으며, 또한 'ㄴ', 'ㄹ', 'ㅁ' 등도 초성일 때와 종성일 때를 분리하였다. 파열음 뒤에 나타나는 기음(aspiration)도 따로 모델링하였다. 모음의 경우, 복모음들도 각기 하나의 모델로 정의하였으며, '외', '왜', '웨' 등과 같이 보통사람이 뚜렷이 구별하여 발성하지 않는 모음들은 하나의 모델로 통합하였다. 마찬가지 이유로 '애', '예'도 하나의 모델로 구성하였다. 마지막으로 유성폐쇄, 무성폐쇄, 그리고 voice offset등을 각기 하나의 모델로 정의하였다. 각 음소를 나타내는 HMM은 기본적으로 그 구조가 그림 1과 같다. 여기서 세 개의 self-loop는 한 음소가 시작(B), 중간(M) 그리고 끝(E)의 세 부분으로 이루어진 것을 나타내며 그림에서 아랫부분의 천이는 4 프레임 이하의 짧은 신호를 모델링한다.

표 1. 유사음소 목록
Table 1. Phoneme-like unit list.

유사음소	음소기호	유사음소	음소기호
/아/	a	/어, 으/	er
/이/	i	/위/	ui
/오/	o	/와/	wa
/우/	u	/왜, 외/	oae
/요/	yo	/ㅊ/	ch1
/애, 에/	ae	/ㅅ/	s1
/여/	yeo	/ㄲ/	gg1
/ㅈ/	j1	/ㅂ/	b1
/ㄷ/	d1	/ㅌ/	t1
/ㄱ/	g1	/ㅍ/	p1
/ㅎ, 기음/	h1	/ㅉ/	jj1
/초성 ㄴ/	n1	/ㅃ/	bb
/종성 ㄴ/	n3	/종성 ㅁ/	m3
/초성 ㅁ/	m1	/초성 ㄹ/	l1
/유성음 ㄷ/	dv	/종성 ㄹ/	l3
/액/	aeg	/종성 ㅇ/	ng3
/옥/	og	/윗/	uis
/앞/	ap	/업/	eop
/ㅋ/	k1	/묵음/	sil
/유성폐쇄/	clv	/무성폐쇄/	clu
/voice offset/	vof		

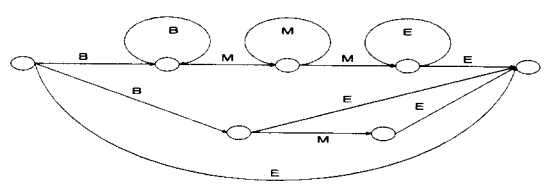

그림 1. 훈련에 사용된 음소모델 기본구조
Fig. 1. Phoneme model base strucure used in training.

인식과정은 다음과 같다. 먼저, 각 단어에 대해 음소의 조합으로 발음사전을 구성한다. 발음사전에 따라 음소 모델을 불러와 각 단어 모델을 만들고 들어온 입력과 Viterbi score를 구한다. 각 단어에 대해서 위의 과정을 반복하여 score가 가장 높은 등록 단어를 인식 단어로 간주한다.

음소가 실제 단어에서는 다양한 형태로 발음될 수 있다. 이것은 인식시 종종 오류의 원인이 되는데, 각 음소가 단어에서 어떤 음소 사이에 있는지에 관한 정보를 HMM에 적용함으로써 인식성능을 향상시킬 수 있다. 즉, 음소 모델에서 각 음소는 문맥에 독립적으로 나타나지만, 실제로 음소란 한 음소의 전후 상황에 의해 많은 영향을 받기 때문에[10] 앞 쪽의 음소와 뒤따라 오는 음소를 구별하여 각각에 대해서 모델을 정의하는 것이다. 앞, 뒤 음소를 다 고려할 경우를 triphone, 한 쪽만을 고려할 경우를 biphone이라 한다. 좌우의 음소를 다 고려한 triphone 모델은 단순한 음소 모델보다 훨씬 정밀하나 모델의 수가 많기 때문에 이들을 잘 학습시키기 어렵다. 이에 비해 biphone은 triphone보다 학습이 쉬울 뿐 아니라 음소의 종속적 특성을 비교적 잘 나타낸다.

본 논문에서는 전체 훈련데이타에 대한 각 음소의 발생빈도비로서 각 biphone에 대한 음소 발생 확률을 정의하여 분할(segmentation)이 어려운 종성을 제외한 초성과 종성에서의 확률을 구하고, 이를 인식실험에 이용하였다. 예를 들면, 음성데이타가 초기 i에서 부분 확률을 가지고 있을 경우, 이런 음성데이타가 중성 j로 넘어갈 때 그 음성데이타가 관측될 확률은 음소쌍의 발생빈도에 따라 증가되어진다. 이때 사용되는 증가확률을 transP(i, j)로 정의하면, 초성 i와 중성 j인 음소쌍의 발생 확률을 matrix의 형태로 나타낼 수 있다. 인식실험에 사용되는 명령어 및 검색어들만을 고려하여 구해진 전체 훈련 데이타에 대한 각 음소쌍의 발생 확률값을 표 2에 나타내었다. 각 음소쌍의 발생 확률 값은 전체 인식어에 대한 각 음소의 발생빈도비로서 정의하여 인식시 확률값에 더해주는 방식을 이용하였다. 아래에 "닫기"라는 음성에 대해 간단한 발음 사전 구조의 예를 보였다. 먼저 부정확한 끝점검출에 의해 포함되는 음성 시작전의 묵음을 모델링하고 각 음소에 해당하는 유사음소모델로 단어를 구성한다. 초성 'ㄷ'은 *d1*, 초성과 중성사이의 '+'는 표 2의 음소쌍 발생확률, 중성 '아'는 a로 정의하여 단어를 구성한다. 특히, 종성 'ㄷ'이 무시되고 초성 'ㄱ'의 격음화와 같은 현상들은 일반적인 발음법에 따른다.

예) 닫기 ; sil d1+a sil gg1+i sil

표 2. 전체 훈련데이타에서 초성과 중성 음소쌍의 발생 확률

Table 2. Probability of Phoneme pair occurrence between initial and middle sounds for training data.

	ㅏ	ㅗ	ㅜ	ㅘ	ㅣ	ㅟ	ㅐ	ㅛ	ㅓ	ㅙ
ㅊ	0.03	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00
ㅅ	0.03	0.03	0.00	0.00	0.00	0.00	0.03	0.00	0.10	0.03
ㄷ	0.07	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00
ㅎ	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
ㄹ	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.07	0.03	0.00
ㅈ	0.00	0.03	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.00
ㅂ	0.03	0.03	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ㅉ	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ㅍ	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00

Ⅲ. 정보검색 시스템의 사용자 인터페이스

본 논문에서 사용된 정보검색 시스템은 비정형화 정보형태인 Full Text Retrieval 시스템으로 미국의 BRS 사의 정보검색 엔진인 BRS/SEARCH를 이용하였고, 음성인식을 위해서는 사운드 카드와 음성 입출력 장치가 갖추어진 PC를 이용하였다. 구현된 사용자 인터페이스는 HMM을 이용한 음성인식 시스템과 정보검색 시스템인 BRS/SEARCH가 결합된 형태로서 시각적 효과를 고려하여 Windows API 함수(Low-Level Sound API 함수 포함)[11]를 이용하여 개발하였다.

1. 시스템의 기능

정보검색 시스템의 사용자 인터페이스는 사용자가 쉽게 이용할 수 있도록 기능을 단순화하고 음성출력을 통해 사용자에게 다음 단계에서 수행해야 할 과정을 설명함으로써 불특정 다수의 일반 사용자가 편리하게 이용하도록 하였다. 특히, 사용자가 인터페이스 화면에서 검색 가능한 DB와 검색리스트를 보면서 선택할 수 있게 하였으며 음성인식을 이용하여 검색어를 입력할 수 있도록 하였다. 인식어 루틴을 linked-list 방법을 이용하여 순차적 인식이 가능하도록 하였다. 그리고 음성인식 루틴과 마우스를 이용한 경우가 동일한

결과를 가지도록 하여 음성과 마우스를 혼용하여 사용할 수 있도록 하였다. 특히 화면 하단에 다음 수행 루틴을 문장으로 표시하고 인식 가능한 단어 리스트를 툴바 아래에 나타내어 사용자의 편리성을 고려하였다. 예를 들면, DB선택 리스트 박스가 팝업되면 "DB를 선택하십시요."라는 문장이 화면 하단에 나타나고 "아래로, 위로, 처음, 마지막, 선택, 취소"와 같은 DB가 선택된 상태에서 인식 가능한 단어 리스트를 툴바 아래에 나타낸다.

구현된 정보검색도구는 하나의 인터페이스상에서 음성인식 시스템과 정보검색 시스템인 BRS가 동시에 존재하는 형태를 가지도록 하여 DB 선택과 검색을 실행하는 인터페이스 화면상에서 사용자가 직접 음성인식 시스템의 메뉴에 접근할 수 있고 음성이 녹음되는 과정을 보여주는 progress 바를 인터페이스 화면에서 직접 보면서 음성 녹음을 제어할 수 있도록 하였다. 이는 음성인식 시스템에 필요한 메뉴와 BRS상에서 필요한 메뉴를 하나의 인터페이스 화면에 모두 포함시킴으로써 가능해진다. 이렇게 함으로써 음성인식 모듈과 BRS사이에서 존재하는 윈도우간의 불필요한 메세지 전달을 없앨 수 있다. 표 3에서 구현된 시스템에서 사용되는 명령어들의 기능을 간단히 설명하였다.

표 3. 명령어의 종류 및 기능
Table 3. List of commands and their functions.

명 령 어	기 능
다운로딩	검색된 결과를 파일로 저장한다.
파일	메뉴바로 이동한다.
종료	인터페이스 프로그램을 종료한다.
DB선택	DB명과 간단한 설명을 리스트 박스의 형태로 디스플레이한다.
DB종료	선택된 DB와 모든 검색 상태를 종료한다.
앞문서	현재 문서의 앞문서를 출력한다.
뒷문서	현재 문서의 뒷문서를 출력한다.
첫문서	검색된 문서중 첫문서를 출력한다.
마지막문서	검색된 문서중 마지막 문서를 출력한다.
목차	인터페이스 화면의 메뉴에 대한 도움말을 제공한다.
사용법	도움말 기능에 대한 설명을 제공한다.
도움말	도움말 기능을 제공한다.
위로	검색결과 화면이나 DB선택 화면에서 한 줄씩 위로 스크롤한다.
아래로	검색결과 화면이나 DB선택 화면에서 한 줄씩 아래로 스크롤한다.
오른쪽	검색결과 화면에서 오른쪽으로 스크롤한다.
왼쪽	검색결과 화면에서 왼쪽으로 스크롤한다.
검색	검색모드로 돌아가 검색어를 기다린다.
선택	DB박스에서 DB를 선택한다.
취소	DB를 선택하지 않거나 이전 모드로 돌아간다.
처음	DB박스 또는 검색결과 화면에서 제일 윗쪽으로 간다.
마지막	DB박스 또는 검색결과 화면에서 제일 아래쪽으로 간다.
페이지 업	검색결과 화면에서 앞 페이지로 스크롤한다.
페이지 다운	검색결과 화면에서 다음 페이지로 스크롤한다.
DB	DB선택과 동일한 기능을 수행한다.
문서	메뉴바중 문서로 이동한다.

2. 음성명령어 및 단어사전 구성

구현된 시스템에서 명령어와 많이 검색되는 검색어들은 발음사전에 미리 정의되어져 있다. 정의된 발음사전은 하나의 파일로 구성되어 있고 인식가능한 명령어와 검색어들은 텍스트 형태로 표현된다. 발음사전에서 사용되는 각 음소모델은 확장자 hmm을 갖는 파일의 형태로 저장되어 있다. 인식가능한 단어들은 표 1에서 설명한 유사음소모델의 결합으로 이루어지며 음소모델의 발음표기에 따라 결합시키게 된다. 표 4는 몇 개의 단어에 대해 음소모델의 결합에 의해 단어사전을 구성한 예를 보인 것이다. 명령어와 검색어의 구분은 등록순서에 따라 이루어지며 만약 새로운 검색어를 등록하고자 할 때에는 반드시 마지막 라인에 등록해야 한다.

표 4. 단어사전 구성 예
Table 4. Example of building word-dictionary.

인식 단어	음 소 모 델
문서 (doc)	m1 u n3 clu s1 er vof
오른쪽(right)	o l1 er n3 clu jj1 og
마지막(last)	m1 a j1 i m1 a g1

일반적으로 인식 시스템을 구현하는데 각 단어의 발음은 일련의 음소단위로써 나타내어진다고 전제하였다. 그러나 실제 발음되는 단어는 사용자에 따라 많은 변화를 보일 수 있다. 따라서 본 연구에서 구현된 시스템에서는 다른 발음이 존재할 수 있는 단어나 유사음소로의 분할이 어려운 단어에 대해서는 몇 개의 단어 모델을 구성하여 하나의 단어로 인식할 수 있게 하였다. 예를 들면, "파일"이라는 단어에 대해서는 "화일", "파일"이라는 발음이 존재할 수 있으므로 아래와 같이 발음사전을 구성하여 동일한 단어로 인식하게끔 구성하였다.

예) file(화일) : h1 wa i l3
file(파일) : p1 wa i l3

사용자가 새로운 검색어를 음성인식 시스템에서 이용하기 위해서는 발음사전에 등록을 해야한다. 등록방법은 단어사전을 이루고 있는 "wtest.dic"라는 파일을 열고 마지막 라인에 원하는 단어를 등록하면 된다. 단어등록은 먼저 인식되는 단어를 적은 후 그 단어가 발음되는 음소모델을 표 2를 참조하여 순서에 따라 간격을 두고 적으면 하나의 단어에 대해 등록이 끝난다. 이때, 주의해야 할 점은 단어가 구성되는 음소들 사이

에 휴지 구간이 존재하는지 반드시 점검해야 한다. 휴지 또는 묵음 구간은 "sil"라는 음소모델명으로 표현한다. 그리고 '스', '트', '츠'와 같은 음절은 하나의 무성음 음소모델로 표현한다. "시스템"이라는 검색어를 등록할 경우의 예를 보이면 다음과 같다.

예) system s1 i s1 sil t1 ae m3

Ⅳ. 실험 및 고찰

1. 음성인식 실험

인식에 사용된 데이타는 8kHz, 16비트의 PCM으로 샘플링된 음성신호로 전달함수가 $1-0.95z^{-1}$인 필터를 통하여 전처리를 한다. 이와 같은 음성 샘플로부터 자기상관방법에 의한 LPC(Linear Predictive Coding) 계수를 구하고, 이들로부터 11개의 LPC cepstrum 계수가 추출되는데, 이 때 한 프레임의 간격은 200 샘플로 하여 100 샘플씩 이동하면서 계산하였다. 시작점, 끝점 검출은 에너지 레벨을 이용하여 불필요한 잡음부분을 제거하였다. 정보검색 시스템에 사용되는 명령어로 선정된 25 개의 단어를 16명의 화자가 각기 5회씩 발성하여 이중 8명의 데이타를 모델의 훈련용으로 사용하고 나머지 8명의 데이타를 인식 실험용으로 사용하였다. 훈련용 데이타를 발성한 화자에 대해 별도로 각 단어를 5회씩 더 발성하게 하여 화자 종속 실험용으로 사용하였다. 훈련용 데이타는 모두 음성파형의 변화를 눈으로 관찰하고 각각의 음소들의 경계점과 각 음소의 이름을 음소단위로 레이블링하였다.

본 연구의 인식실험에 사용되는 음소모델은 전부 41개이며, 이들에 대한 각 HMM 모델은 전향, 후향 알고리즘을 사용하여 실제 음성을 사용하여 학습된다. HMM 모델은 SPHINX 모델을 사용하였으며, 각 state의 mixture를 위한 observation probability density function의 수는 7개이다. 이 경우 초기의 파라미터가 잘 맞춰진 경우에는 2회 정도의 알고리즘 수행으로도 충분하나 본 연구에서는 좀 더 정확한 모델 파라미터를 얻기 위하여 최대 15회의 반복을 허용하였다. 이와 같이 학습을 통하여 계산된 천이 확률은 음성인식 과정에서 그대로 사용되지만 출력확률이 극히 작아지는 경우를 피하기 위해 floor smoothing을 이용하여 smoothing 과정을 거치게 된다.

표 5. 화자종속 인식실험의 경우 오인식률
Table 5. Error rate in the case of speaker-dependent.

단어 \ 화자	화자 1	화자 2	화자 3	화자 4	화자 5	화자 6	화자 7	화자 8	단어별
파일	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/40
문서	0/5	0/5	1/5	2/5	1/5	0/5	0/5	0/5	4/40
검색	1/5	0/5	1/5	0/5	1/5	0/5	1/5	1/5	5/40
종료	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/40
다운로딩	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/40
위로	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/40
아래로	1/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	1/40
오른쪽	0/5	0/5	0/5	0/5	0/5	0/5	3/5	2/5	5/40
왼쪽	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/40
취소	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/40
페이지 업	0/5	1/5	0/5	0/5	0/5	0/5	0/5	0/5	1/40
페이지 다운	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/40
사용법	2/5	0/5	0/5	0/5	1/5	1/5	0/5	0/5	4/40
첫문서	0/5	2/5	2/5	0/5	0/5	2/5	0/5	0/5	6/40
마지막문서	0/5	0/5	0/5	0/5	0/5	0/5	2/5	0/5	2/40
앞문서	0/5	2/5	0/5	1/5	1/5	0/5	1/5	0/5	5/40
뒷문서	3/5	0/5	1/5	1/5	2/5	0/5	0/5	0/5	7/40
도움말	1/5	0/5	1/5	0/5	1/5	0/5	1/5	1/5	5/40
목차	0/5	1/5	2/5	2/5	2/5	2/5	0/5	0/5	9/40
처음	1/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	1/40
마지막	1/5	0/5	0/5	0/5	0/5	0/5	3/5	2/5	6/40
DB	0/5	1/5	0/5	0/5	0/5	0/5	0/5	0/5	1/40
DB선택	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/40
DB종료	2/5	0/5	0/5	0/5	1/5	1/5	0/5	0/5	4/40
선택	0/5	0/5	0/5	0/5	0/5	0/5	2/5	0/5	2/40
화자별	12/125	7/125	8/125	6/125	10/125	6/125	13/125	6/125	68/1000

표 6. 화자독립 인식실험의 경우 오인식률
Table 6. Error rate in the case of speaker-independent.

단어 \ 화자	화자9	화자10	화자11	화자12	화자13	화자14	화자15	화자16	단어별
파일	0/5	0/5	1/5	0/5	0/5	0/5	0/5	2/5	3/40
문서	0/5	2/5	0/5	1/5	0/5	0/5	0/5	1/5	4/40
검색	1/5	0/5	0/5	3/5	1/5	2/5	1/5	0/5	8/40
종료	0/5	0/5	0/5	0/5	0/5	0/5	0/5	1/5	1/40
다운로딩	0/5	1/5	0/5	0/5	1/5	0/5	0/5	0/5	2/40
위로	0/5	0/5	0/5	1/5	0/5	0/5	0/5	0/5	1/40
아래로	0/5	0/5	2/5	1/5	0/5	1/5	0/5	0/5	4/40
오른쪽	1/5	0/5	0/5	0/5	0/5	0/5	1/5	2/5	4/40
왼쪽	0/5	2/5	1/5	0/5	3/5	0/5	0/5	1/5	7/40
취소	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/40
페이지 업	0/5	1/5	0/5	0/5	0/5	0/5	2/5	1/5	4/40
페이지 다운	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/40
사용법	0/5	0/5	1/5	0/5	0/5	3/5	0/5	0/5	4/40
첫문서	2/5	0/5	0/5	0/5	0/5	0/5	1/5	1/5	4/40
마지막문서	0/5	1/5	1/5	0/5	0/5	0/5	0/5	0/5	2/40
앞문서	1/5	1/5	0/5	0/5	2/5	1/5	0/5	0/5	5/40
뒷문서	0/5	0/5	1/5	0/5	1/5	0/5	1/5	1/5	4/40
도움말	1/5	0/5	0/5	0/5	1/5	0/5	1/5	0/5	3/40
목차	0/5	0/5	3/5	2/5	2/5	2/5	0/5	0/5	9/40
처음	0/5	1/5	0/5	2/5	0/5	0/5	0/5	0/5	3/40
마지막	1/5	0/5	0/5	1/5	0/5	0/5	0/5	0/5	2/40
DB	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/40
DB선택	2/5	0/5	1/5	2/5	1/5	1/5	0/5	0/5	7/40
DB종료	0/5	3/5	0/5	0/5	0/5	2/5	0/5	0/5	5/40
선택	3/5	0/5	1/5	1/5	2/5	0/5	0/5	0/5	7/40
화자별	12/125	12/125	12/125	14/125	14/125	12/125	7/125	10/125	93/1000

각 음소모델의 훈련은 먼저 모델 파라미터가 global optimum에 접근할 수 있도록 간단한 균일분포를 가정하고 모델 초기화를 수행한 후 Baum-Welch 알고리즘의 반복수행을 통하여 각 모델 파라미터가 결정된다. 파라미터의 재추정시 underflow를 방지하기 위해 전향-후향확률을 균일화(normalization)하는 방식으로 scaling을 하였고 불충분한 훈련데이타를 보완하기 위하여 관측확률을 10^{-4}으로 floor smoothing하였다.

표 5 및 6에 화자종속 및 화자독립의 경우에 대한 인식실험 결과를 나타내었다. 화자종속실험에서는 6.8%, 화자독립 실험에서는 9.3%의 오인식률을 보였다. 단어별로는 음소 '어'가 포함되는 단어, '문서, 검색, 오른쪽, 첫문서, 뒷문서' 등에서 오인식이 많았다.

2. 구현된 사용자 인터페이스

개발된 정보검색도구는 하나의 인터페이스상에서 음성인식 시스템과 정보검색 시스템인 BRS가 동시에 존재하는 형태이다. 그림 2는 구현된 시스템을 실제로 실행시켰을 때의 화면을 보인 것이다.

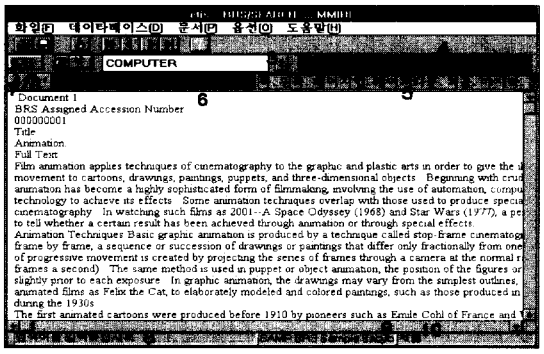

그림 2. 구현된 인터페이스의 전체화면
Fig. 2. Display of the implemented interface.

그림 2에서 각 번호에 해당하는 기능은 다음과 같다.

- 1 : DB선택 버튼으로서 이 버튼을 누르면 DB 리스트 박스가 팝업된다. 여기에서 사용자가 검색하고자 하는 DB를 선택한다.
- 2 : 검색 시작 버튼으로서 이 버튼을 누르면 입력된 검색어를 검색한다. 그리고 검색명령에 의해 선택된 문서가 화면에 출력된다.
- 3 : 검색어 입력 에디트 박스와 버튼으로서 에디트 박스에는 키보드를 이용하여 검색어를 입력할 수 있고 버튼을 누르면 검색 가능한 리스트박스가 팝업된다. 또한, 입력된 검색어외에 사용자가 많이 사용하는 검색어 리스트를 미리 만들어 놓고 음성으로 이를 검색할 수도 있다.
- 4 : 현재 시간을 표시
- 5 : 현재의 상태에서 인식가능한 명령어단어들의 리스트를 보여준다. 사용자는 여기에 표시되는 단어에 대해서는 음성으로 입력이 가능하다.
- 6 : 마이크 입력 레벨 표시
- 7 : 마이크 ON/OFF 기능 버튼. 음성명령을 사용할 경우 사용자는 이 버튼을 활성화 시켜야 한다.
- 8 : 사용자에게 현재 단계에서 수행해야 할 과정을 음성으로 들려주면서 문장으로 출력한다. 이 기능은 특히 검색 시스템에 익숙하지 않은 사용자가 음성출력을 듣고 다음 단계로 쉽게 넘어갈 수 있게 설명해준다.
- 9 : 현재 사용중인 DB명을 표기한다.
- 10 : 현재 음성 명령에 의해 인식된 결과를 보여준다.

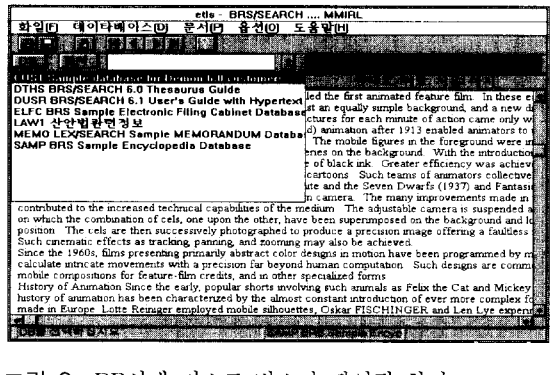

그림 3. DB선택 리스트 박스가 팝업된 화면
Fig. 3. Popup screen of DB list box.

구현된 음성인식 인터페이스에서는 화면이 최초로 팝업되면서 간단한 로그인 박스가 활성화되고 사용자는 이름과 암호를 입력하여 로그인 과정을 수행한다. 로그인 과정이 수행되고 나면 "DB를 선택하십시요." 라는 음성 메세지가 들려오고 DB를 선택할 수 있는 리스트 박스가 팝업된다. 이 때 박스안의 첫 번째 DB가 활성화되고 음성이나 마우스를 이용하여 원하는 DB를 선택할 수 있다. 만약 원하는 검색 과정을 모두 수행한 후 새로운 DB를 선택하고자 할 경우에는 마

우스를 이용하여 메뉴에서 "데이타베이스 선택"이나 툴바에 있는 DB버튼을 누르거나 검색모드 환경에서 "DB"라는 음성을 이용하여 DB 리스트 박스를 활성화할 수도 있다. 원하는 DB로의 이동은 "아래로", "위로", "처음", "마지막"같은 간단한 음성 명령을 사용할 수 있으며 "선택", "취소"라는 음성명령을 이용하여 원하는 DB를 선택하거나 DB선택을 취소할 수 있다. DB가 선택되면 DB명이 화면 하단의 중앙에 표기된다. 그림 3은 검색모드 환경에서 새로운 DB를 선택하기 위해 리스트 박스를 음성을 이용하여 팝업시킨 화면이다.

DB선택이 이루어지면 선택된 DB가 인터페이스 화면에 디스플레이 되면서 "검색어를 입력하십시요."라는 간단한 음성 메세지가 들리고 동일한 문장이 화면 왼쪽 하단에 표기되어진다. 사용자는 원하는 검색어를 에디트 박스에 키보드를 이용하여 직접 입력하거나 미리 선정된 검색어들 중에서 음성을 이용하여 입력할 수 있다. 미리 선정된 검색어 리스트는 에디트 박스의 버튼을 누르면 사용자가 볼 수 있다. 만약 원하는 단어가 리스트에 없는 경우 발음단어사전에 등록하여 다음 검색시에는 이용이 가능해진다. 음성을 이용하여 검색어가 입력되면 시스템에서 자동적으로 그 단어를 검색하여 검색된 결과를 인터페이스 화면에 디스플레이한다. 본 연구에서 구현된 인터페이스는 검색 기능을 위주로 사용자가 간단하고 편리하게 사용할 수 있도록 복잡한 메뉴의 구성을 피하고 사용자에게 기본적으로 필요한 기능들을 메뉴화하였다.

V. 결 론

본 연구에서는 일반 사용자가 편리하게 정보검색을 할 수 있도록 정보검색 도구의 입력기능, 검색 및 갱신 기능 등을 음성을 이용하여 수행할 수 있는 사용자 인터페이스를 구현하였다. 구현된 시스템은 크게 음성 인터페이스를 위한 음성인식 시스템과 정보검색도구의 사용자 인터페이스로 나누어진다. 음성인식 시스템의 경우 HMM(Hidden Markov Model)을 이용한 화자 독립적인 인식 시스템을 구현하였는데, 인식실험 결과, 16 명의 남성화자를 대상으로 실시한 화자종속, 화자 독립 인식실험에서 각각 93.2 %, 90.7 %의 인식률을 얻을 수 있었다.

개발된 정보검색 도구의 인터페이스는 음성인식에 의한 정보검색에 적합하도록 기능을 단순화하여 불특정 다수의 일반 사용자가 쉽게 이용하도록 하였다. 특히, 사용자가 인터페이스 화면에서 검색 가능한 DB와 검색리스트를 보면서 선택할 수 있게 하였으며 음성인식을 이용하여 검색어를 입력할 수 있도록 하였다. 인식어 루틴을 linked-list 방법을 이용하여 순차적 인식이 가능하도록 하였다. 그리고 음성인식 루틴과 마우스를 이용한 경우가 동일한 결과를 가지도록 하여 음성과 마우스를 혼용하여 사용할 수 있도록 하였다.

참 고 문 헌

[1] Christopher Schmandt, *Voice Communication with Computers*, Van Norstrand Reinhold, New York, 1994.

[2] Jeffrey Richter, *Advanced Windows*, Microsoft Press, 1995.

[3] F. Jelinek, "Continuous Speech Recognition by Statistical Method", *Proc. IEEE*, Vol. 64, no. 4, April 1976.

[4] L. R. Bahl, F. Jelinek, B. L. Lewis, and R. Mercer, "A Maximim Likelihood Approach to Continuous Speech Recognition", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. PAMI-5, pp. 179-190, March 1983.

[5] J. K. Baker, "The DRAGON System - An Overview", *IEEE Tran. ASSP*, Vol 23, no. 1, Feb. 1975.

[6] K-F. Lee, H. W. Hon, and R. Reddy, "An Overview SPHINX speech recognition system", *IEEE Trans. ASSP*, Vol. 38, pp. 35-45, Jan. 1990.

[7] L. R. Rabiner, J. G. Wilpon, and F. K. Soong, "High Performance Connected Digit Recognition Using Hidden Markov models", *IEEE ICASSP*, Apr. 1988.

[8] L. R. Rabiner and S. E. Levinson, "A Speaker-Independent, Syntax-Directed, Connected Word Recognition System Based on Hidden Markov Models and Level Building", *IEEE Trans. ASSP*, Vol. 33, no. 3, pp. 561-573, June. 1985

[9] A. H. Gray and Jr., J. D. Markel, "Distance Measures for Speech Process-

ing," *IEEE Trans. Acoust, Speech, Signal Processing*, Vol. ASSP-24, No. 5, pp. 380-391, Oct. 1976.

[10] A. Rudnicky, L. Baumeister, K. DeGraaf, and E. Lehmann, "The Lexical Access Component of the CMU Continuous Speech Recognition System", *IEEE ICASSP*, April 1987.

[11] Microsoft, *Windows Multimedia Programmer's Workbook*, Microsoft Press.

저 자 소 개

金 正 哲(準會員)
1995년 2월 경북대학교 전자공학과(공학사). 1998년 2월 경북대학교 전자공학과(공학석사). 주관심분야는 디지털 신호처리, 음성신호처리, 음성인식

裵 建 星(正會員)
1977년 2월 서울대학교 전자공학과(공학사). 1979년 2월 한국과학기술원 전기 및 전자공학과(공학석사). 1989년 5월 University of Florida(공학박사). 1979년 3월~현재 경북대학교 전자공학과 교수. 주관심분야는 음성분석 및 인식, 디지털 신호처리, 디지털 통신, 음성 부호화, 웨이브렛 분석 등