

한국어 음성인식 플랫폼 (ECHOS) 개발

Development of a Korean Speech Recognition Platform (ECHOS)

권 오 욱*, 권 석 봉**, 장 규 철***, 윤 성 락***, 김 용 래*, 장 광 동*, 김 희 린**
유 창 동***, 김 봉 완****, 이 용 주****

(Oh-Wook Kwon*, Sukbong Kwon**, Gyucheol Jang***, Sungrack Yun***, Yong-Rae Kim*,
Kwang-Dong Jang*, Hoi-Rin Kim**, Changdong Yoo***, Bong-Wan Kim****, Yong-Ju Lee****)

*충북대학교, **한국정보통신대학교, ***한국과학기술원, ****음성정보기술산업지원센터
(접수일자: 2005년 8월 12일; 채택일자: 2005년 9월 20일)

교육 및 연구 목적을 위하여 개발된 한국어 음성인식 플랫폼인 ECHOS를 소개한다. 음성인식을 위한 기본 모듈을 제공하는 ECHOS는 이해하기 쉽고 간단한 객체지향 구조를 가지며, 표준 템플릿 라이브러리 (STL)를 이용한 C++ 언어로 구현되었다. 입력은 8 또는 16 kHz로 샘플링된 디지털 음성 데이터이며, 출력은 1-best 인식결과, N-best 인식결과 및 word graph이다. ECHOS는 MFCC와 PLP 특징추출, HMM에 기반한 음향모델, n-gram 언어모델, 유한 상태망 (FSN)과 렉시컬트리를 지원하는 탐색알고리즘으로 구성되며, 고립단어인식으로부터 대어휘 연속음성인식에 이르는 다양한 태스크를 처리할 수 있다. 플랫폼의 동작을 검증하기 위하여 ECHOS와 hidden Markov model toolkit (HTK)의 성능을 비교한다. ECHOS는 FSN 명령어 인식 태스크에서 HTK와 거의 비슷한 인식률을 나타내고 인식시간은 객체지향 구현 때문에 약 2배 정도 증가한다. 8000단어 연속음성인식에서는 HTK와 달리 렉시컬트리 탐색 알고리즘을 사용함으로써 단어오류율은 40% 증가하나 인식시간은 0.5배로 감소한다.

핵심용어: 음성인식, 탐색알고리즘

투고분야: 음성처리 분야 (2.3)

We introduce a Korean speech recognition platform (ECHOS) developed for education and research purposes. ECHOS lowers the entry barrier to speech recognition research and can be used as a reference engine by providing elementary speech recognition modules. It has an easy simple object-oriented architecture, implemented in the C++ language with the standard template library. The input of the ECHOS is digital speech data sampled at 8 or 16 kHz. Its output is the 1-best recognition result, N-best recognition results, and a word graph. The recognition engine is composed of MFCC/PLP feature extraction, HMM-based acoustic modeling, n-gram language modeling, finite state network (FSN)- and lexical tree-based search algorithms. It can handle various tasks from isolated word recognition to large vocabulary continuous speech recognition. We compare the performance of ECHOS and hidden Markov model toolkit (HTK) for validation. In an FSN-based task, ECHOS shows similar word accuracy while the recognition time is doubled because of object-oriented implementation. For a 8000-word continuous speech recognition task, using the lexical tree search algorithm different from the algorithm used in HTK, it increases the word error rate by 40% relatively but reduces the recognition time to half.

Keywords: Speech recognition, search algorithm

ASK subject classification: Speech Signal Processing (2.3)

1. 서론

음성인식기의 상용화 추세에 따라서 음성인식의 세부 기술 및 구현에 관심이 있는 사람들이 늘어나고 있다.

그러나 기존에 공개된 음성인식기는 구현이 복잡하고 문서가 충분하지 않아서 내부구조를 수정할 필요가 있는 자신의 아이디어를 연구자들이 실험하는데 어려움이 있다. 이에 본 논문에서는 교육 및 연구 목적을 위하여 쉽게 이해할 수 있고 설명서를 가진 한국어 음성인식 플랫폼을 개발하였다. 이러한 음성인식 플랫폼은 새로운 아이디어를 가진 연구자의 음성인식 진입 장벽을 낮추고,

연구역량을 핵심기술에만 집중할 수 있도록 하며, 대학 및 연구소에서 발표하는 음성인식 연구결과와 검증에 활용될 수 있다.

음성인식 분야의 진입장벽 감소 및 저변 확대를 위하여 외국에서는 대학을 중심으로 음성인식 플랫폼을 공개하고 있다. 사실상의 표준 음성 데이터베이스와 공개 소프트웨어를 제공함으로써 새로운 아이디어를 쉽게 검증할 수 있도록 하고 있다. 기존에 공개된 음성인식 플랫폼으로서 HTK (Hidden Markov Toolkit)[1], Sphinx [2], Mississippi 대학 음성인식기[3], Julius[4], ezCSR [5] 등이 있다.

HTK[1]은 영국 캠브리지 대학에서 개발된 것으로 음성인식기의 훈련 및 테스트에 사용되는 사실상의 표준 플랫폼이다. C언어로 구현되어 있으며, NIST의 벤치마크 테스트에서도 우수한 성능을 보였다. 최근에는 ETSI의 Aurora-2 및 Aurora-3 프로젝트에서의 기준 인식기로 사용되었다. 현재는 중규모 연속음성인식을 지원하나, 2005년대로 대어휘 연속음성인식을 지원할 예정이다. Sphinx[2]는 C언어로 구현되었으며, 카네기멜론대학교 (CMU)의 자원관리 (Resource management) 테스트에 사용되어 최초로 성공적인 연속음성인식 기술을 선보였다. Mississippi 대학에서 만든 음성인식기[3]는 foundation class의 형태로 개발되어 일반적인 음성처리에 응용될 수 있도록 되어 있다. C++언어로 구현되었으며, 객체기반 연속음성 인식기의 훈련 및 인식을 위한 도구이다. European Telecommunications Standards Institute (ETSI)의 Aurora-4 (5000단어 WSJ) 프로젝트의 기준 인식기로 사용되었다. 일본에서는 1996년부터 교토대를 중심으로 Julius [4]라는 대어휘 연속음성 인식기를 개발하였다. ezCSR[5]은 국내에서 개발된 객체지향 음성인식기로서, C++언어로 구현되었으며, 탐색 트리 (search tree) 및 유한상태망 (finite state network) 기반의 탐색 알고리즘을 지원하며, 한글 발음 사전 생성 및 잡음제거 기능을 지원한다.

ECHOS (Easy Compact Hangeul Object-oriented Speech recognizer)[6]는 ezCSR의 소스 및 소프트웨어 구조를 바탕으로, 오류 수정, 기능 추가, 매뉴얼 보완 과정을 거쳐서 개발된 한국어 음성인식 플랫폼이다. 추가된 기능은 잡음제거, ETSI 특징추출, n-best 탐색, word graph 생성 등이다. ECHOS는 쉽고 작으면서 한글 처리가 가능한 객체기반의 구조를 가지며, 표준 템플릿 라이브러리를 이용한 C++언어로 구현되었다. 개발된

음성인식 플랫폼의 성능을 검증하기 위하여 네트워크 기반의 탐색, 렉시컬 트리 탐색의 2가지 태스크에 대하여 인식을 및 인식시간을 비교하였다.

제2절에서는 ECHOS의 특징, 구조, 기능, 입출력 규격, 응용 프로그램 인터페이스를 기술하였다. 제3절에서는 클래스 다이어그램 및 시퀀스 다이어그램을 이용하여 ECHOS의 소프트웨어 구조 및 동작을 설명하였다. 제4절에서는 개발된 음성인식 플랫폼의 정확성 및 성능을 검증하기 위하여 널리 사용되는 HTK와 인식을 및 인식시간을 비교하였다. 제5절에서 결론을 맺는다.

II. ECHOS

2.1. 특징

ECHOS는 쉽고 문서화가 잘되어 초보자도 쉽게 이해할 수 있는 구조를 갖는다. 고수준의 표준 라이브러리인 standard template library (STL)[7]를 사용하여 프로그램의 가독성을 높이고 본질적인 알고리즘 구현에 집중할 수 있게 한다. 한글 발음 사전, 형태소 분석에 기반한 한글 발음변환 기능을 제공하여 한국어 처리에 불편함을 최소화한다. 사용자가 쉽게 자신의 알고리즘을 치환하여 검증할 수 있도록 객체 지향의 프로그램 구조를 갖는다. 플랫폼의 일부 모듈을 다른 용도에 쉽게 적용할 수 있도록 각 모듈의 사용 예제 프로그램을 제공한다.

잘 갖추어진 문서 및 설명서는 플랫폼이 널리 사용되기 위한 필수 요소이다. 이를 위하여 객체지향 언어로 구현된 소프트웨어의 문서화에 적합한 unified modeling language (UML)[8]을 수정하여, 플랫폼의 구조 및 동작을 기술한다. 그림 1은 이러한 ECHOS의 특징을 요약한다.

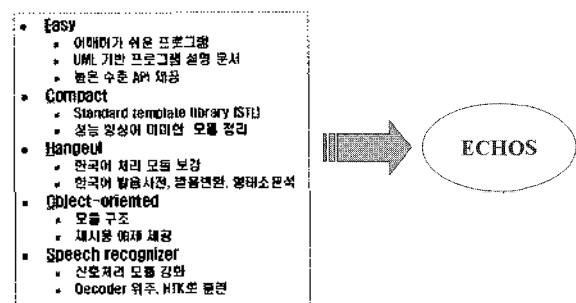


그림 1. ECHOS의 특징
Fig. 1. Features of ECHOS.

2.2. 입출력 규격

입력: Windows에서는 유/무선 마이크 또는 파일을 통하여 입력신호를 받을 수 있으며, Linux 또는 Unix에서는 파일을 통해서만 입력신호를 받는다. 8/16 kHz 샘플링 주파수와 16 비트 PCM 및 8비트 μ -law 포맷을 지원한다.

출력: ECHOS는 1-best, n-best, word graph의 3가지 형태의 인식결과를 제공한다. N-best 인식결과는 word graph 결과로부터 backward tracking을 통하여 얻는다. 인식결과에 대하여 경계 및 likelihood 관련 정보를 제공한다. 순방향 탐색과정으로부터 단어 단위의 경계 (segmentation)정보와 likelihood에 대한 정보를 얻는다. 음소 및 상태 단위의 분할 정보와 likelihood 정보는 1-best 인식결과와 음성신호를 후처리에서 비터비 정렬 (Viterbi alignment)함으로써 얻는다.

2.3. 구조 및 기능

한국어 음성인식 플랫폼은 고립 단어 인식, 연속음성인식, 음성 분할 기능을 수행할 수 있다. 사운드카드로부터 직접 입력되는 음성을 인식하는 온라인 모드와 파일에 저장된 음성을 인식하는 오프라인 모드를 지원한다. 응용 프로그램 개발을 위한 라이브러리, 음성인식 실험을 위한 도구, 음성 파일의 음소단위 분할 도구로서 활용 가능하다.

플랫폼은 그림 2와 같이 신호처리, 음성검출, 특징추출, 음향모델, 발음사전, 언어모델, 탐색, 후처리 모듈을 가진다. 각 모듈의 기능 및 규격은 다음과 같다.

잡음제거: 강인한 음성인식을 위하여 입력신호로부터 배경잡음과 채널잡음을 제거한다. ECHOS는 Spectral subtraction, Wiener filtering 방법을 지원한다.

음성검출: 입력신호로부터 음성 부분만을 찾아내어 음성인식기로 전달한다. ECHOS는 에너지 기반 음성 검출 알고리즘을 제공한다.

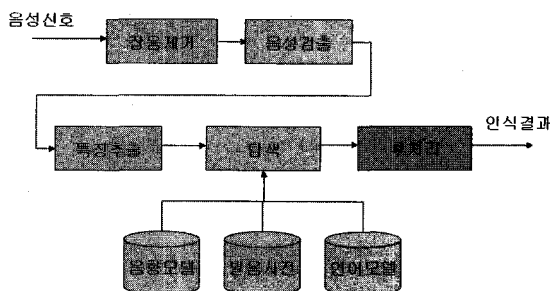


그림 2. ECHOS의 구조

Fig. 2. Block diagram of ECHOS.

특징추출: 입력신호로부터 음성인식에 유용한 특징을 추출한다. 현재 버전에서 mel frequency cepstral coefficient (MFCC)[11], perceptually linear prediction (PLP) 계수[12] 특징을 지원하며, 차기 버전에서는 ETSI 특징[13]을 추가할 예정이다.

음향모델: 음향모델은 음향특징을 모델링하여 음향단위에 대한 likelihood을 계산하는 모듈이다. ECHOS는 연속HMM[9,10]을 채택하며, diagonal 또는 full 공분산 행렬을 지원한다. 음향모델은 HTK 포맷을 따른다. 형성된 음향모델에 대해 훈련된 음향모델이 부족하거나 유사한 확률분포를 가지는 상태의 파라미터를 공유 (state-sharing) 기능을 지원한다. Decision tree를 사용하여 음향모델을 갱신하거나 또는 탐색 과정에서 음향모델을 선택할 수 있다.

발음사전: 인식대상 어휘의 발음을 제공하는 모듈이다. ECHOS는 한글처리를 위하여 발음사전에 한글표제어를 사용할 수 있다. 한글 어휘에 대한 자동적으로 발음기호로 표현해 주는 발음사전 생성기를 제공한다. 한 단어는 여러 개의 발음이 가능하도록 다중발음을 지원한다.

언어모델: 언어모델은 연속음성인식을 위하여 문법을 모델링하며, 유한상태망 (FSN)과 통계적 언어모델로 크게 구분된다. FSN은 인식하고자 하는 단어들의 연결관계를 네트워크로 표현하는 것으로서 자유도가 낮기 때문에 인식단어 수가 적은 태스크에 주로 사용된다.

통계적인 언어모델은 단어가 인접하여 발생할 확률을 나타내는 것으로서, 이전 단어의 개수에 따라서 유니그램, 바이그램, 트라이그램으로 나누어진다. 유니그램은 단어 자체의 발생확률을 나타낸 문법이며, 바이그램은 과거의 한 단어로부터 다음에 나타날 단어의 확률을 정의한다. 트라이그램은 과거의 두 단어로부터 다음에 나타날 단어의 확률로 정의되며, 바이그램보다 정교하여 인식성능을 높일 수 있으나 연산량이 많고 복잡하다.

ECHOS는 소규모 태스크를 위한 FSN과 대어휘 연속음성인식을 위한 통계적 언어모델인 바이그램과 트라이그램을 지원한다.

탐색: 탐색모듈은 발음사전 및 언어모델로부터 구성된 탐색네트워크에서 최대 확률을 나타내는 단어열을 찾는 알고리즘이다. 탐색 알고리즘은 크게 소규모 태스크를 위한 FSN 탐색과 대어휘 연속음성인식을 위한 렉시컬트리 탐색으로 나누어진다. FSN 탐색은 인식대상 어휘가 작거나 인식속도가 느리더라도 인식성능에 중점을 두었을 때 사용되는 방식이다. 렉시컬트리 탐색은 인식대상 어

휘가 많거나 인식성능은 떨어지더라도 빠른 인식속도가 요구되는 태스크에 주로 사용된다.

ECHOS는 FSN 탐색과 트리 기반 탐색 알고리즘을 모두 지원한다. 인식을 향상을 위하여 2단계 탐색 기법을 지원한다. 먼저 1단계에서 바이그램을 이용하여 탐색 네트워크를 탐색하고 2단계에서 트라이그램을 이용한 스택 디코딩을 통하여 보다 정확한 인식결과를 얻는다.

후처리: 현재 버전에서는 음소 및 상태 단위의 분할 정보와 likelihood 정보를 얻기 위하여 1-best 인식결과와 음성신호를 비트비 정렬하는 기능을 갖는다. 사용자가 ECHOS의 인식결과를 이용하여 성능 향상을 하고자 하는 경우에 활용될 수 있다.

2.4. EAPI

ECHOS 응용 프로그램 인터페이스(EAPI)는 ECHOS와 응용프로그램의 인터페이스를 나타낸다. EAPI 규격은 사용자 수준에 따라서 두 단계로 제공된다. EAPI는 인식 플랫폼을 제어하거나, 모듈이 개별적으로 사용될 수 있는 인터페이스를 제공한다. 그림 3은 EAPI의 구조를 나타낸다. EAPI Runtime module은 인식엔진을 제어하거나 관련 정보를 접근하고 개별 모듈에 대한 인터페이스를 제공하기 위하여 인식엔진에 덧붙여진 것이다.

III. 구현

3.1. 클래스 다이어그램

ECHOS는 그림 4와 같은 클래스로 구성된다. Speech-Recog 클래스는 사용자 프로그램과의 인터페이스를 담당한다. SearchBase 클래스는 탐색 모듈, 인식에 필요한 모든 모듈을 관리하고, 탐색 알고리즘에 따라서 해당하는 탐색객체를 호출한다. SearchTree 클래스는 대어휘를 위한 lexical 트리를 구성하여 탐색한다. Search-

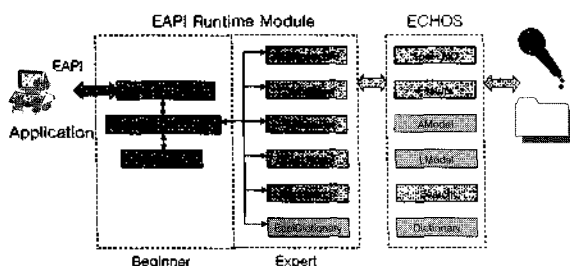


그림 3. EAPI 구조
Fig. 3. EAPI structure.

Network 클래스는 소규모 또는 중규모 어휘를 갖는 음성인식을 위하여 FSN을 구성하여 탐색한다. AudioIO 클래스는 사운드카드 또는 파일로부터 음성을 읽어 들인다. 대략적인 끝점검출 기능도 동시에 수행된다. Feature 클래스는 입력신호로부터 잡음을 제거하고 특징을 추출한다. 정교한 끝점검출이 사용되어 탐색모듈에게 음성입력 완료를 알려준다. AModel 클래스는 음향모델을 읽어 들이고, 입력된 특징에 대한 로그확률을 계산한다. LModel 클래스는 언어모델을 읽어들이고, 이전의 단어 열이 주어질 때 현재단어의 로그확률을 계산한다. Post-Proc 클래스는 인식결과에 다른 지식원을 부가하여 향상된 인식결과를 제공한다.

3.2. 소프트웨어 동작

ECHOS는 그림 5와 같이 동작한다. SpeechRecog 모듈은 사용자 응용 프로그램으로부터의 요구사항에 따라서 적절히 탐색모듈을 호출한다. 탐색모듈은 프레임 단위로 관련 모듈을 순차적으로 호출한다. 탐색모듈은 오

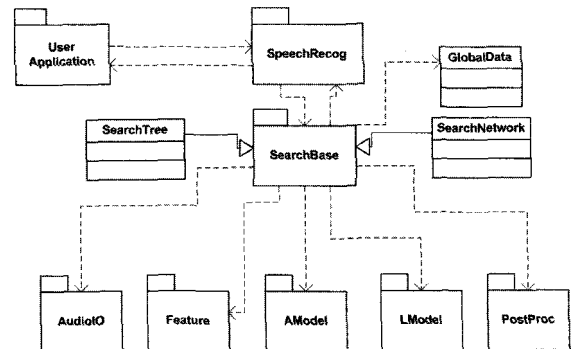


그림 4. ECHOS의 소프트웨어 구조
Fig. 4. Software architecture of ECHOS.

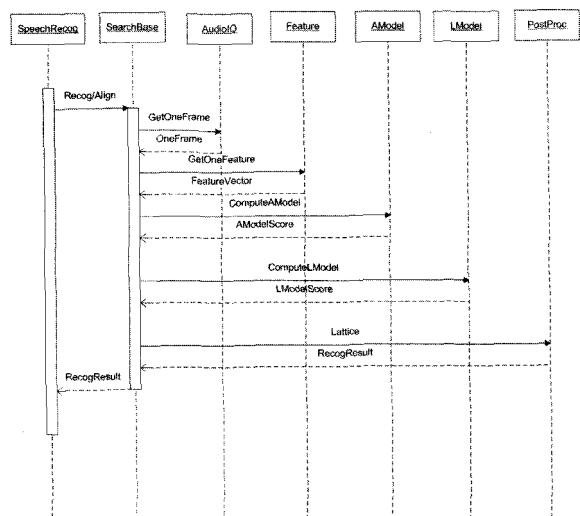


그림 5. ECHOS의 시퀀스 다이어그램
Fig. 5. Sequence diagram of ECHOS.

표 1. FSN 탐색에서의 인식률(%) 비교 실험 결과
Table 1. Evaluation results for FSN search.

발간격	HTK	ECHOS with NR	ECHOS without NR
100	97.5	90.0	97.5
250	97.5	95.0	97.5
500	97.5	95.0	97.5
1000	97.5	95.0	97.5

디오입출력 모듈로부터 한 프레임씩 읽어들이며 특징추출 모듈로 전달한다. 특징추출 모듈에 특징추출로부터 한 프레임씩 읽어들이며 탐색트리 또는 FSN에서 한 프레임씩 전진한다. 끝점이 검출되면 backtracking을 수행하여 word graph 또는 1-best 출력을 얻는다. Search 모듈은 word graph를 후처리 모듈에 전달하여 최종결과를 받는다.

IV. 검증

4.1. FSN 인식

FSN을 이용한 음성 인식 시스템의 성능 검증을 위하여 다음과 같은 소규모 단어 네트워크를 구성하였다.

특징벡터로는 12차 MFCC, 12차 델타 MFCC 및 델타 에너지를 이용한 총 25차 MFCC 특징벡터를 이용하였다. 총 10개의 녹음된 문장에 대하여 테스트를 실시하였으며 잡음제거(NR) 기능을 사용하는 경우와 사용하지 않은 경우로 나누어 실험하였다. 실험을 위한 파라미터는 단어 삽입 패널티 0, 단어 기각 임계값 0으로 HTK와 ECHOS에 동일하게 설정하였다.

비록 작은 규모의 실험이지만 동일한 파라미터설정일

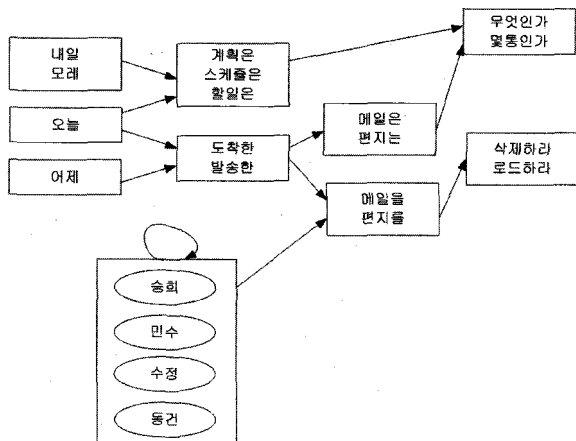


그림 6. 검증 실험을 위한 소규모 FSN
Fig. 6. A small FSN for validation.

표 2. ECHOS와 HTK에서의 연속음성인식 테스트의 인식률(%) 및 인식시간 (HTK에서의 인식시간을 1로 함)
Table 2. Word accuracy (%) and recognition time for continuous speech recognition (recognition time of HTK is assumed to be 1).

탐색 알고리즘	HTK		ECHOS	
	인식률	시간	인식률	시간
플랫렉시콘	85.3	1	85.8	2.5
렉시컬트리	X	X	76.5	0.5

때 동일한 인식 성능을 얻을 수 있었으며, 잡음 제거 기능을 사용하였을 경우에 약간 성능이 떨어지는 이유는 테스트 데이터가 거의 잡음이 없는 환경에서 녹음되어, 잡음 제거 알고리즘 자체의 왜곡이 인식성능에 영향을 끼친 것으로 볼 수 있다.

4.2. 8000단어 연속음성인식

ECHOS의 연속음성인식 검증을 위해 SITEC의 Dict01 (낭독문장 음성 DB)로부터 훈련된 음향모델과 바이그램 언어모델을 사용하였고, 8,670개의 단어로 구성된 발음 사전과 훈련에 사용되지 않은 105개 발화를 테스트에 사용하였다.

HTK에서는 렉시컬트리를 생성할 수 없기 때문에 HTK와 인식성능 비교분석을 위해 ECHOS의 플랫렉시콘과 같은 조건에서 인식실험을 하였다. ECHOS에서는 HTK와 달리 렉시컬트리 탐색을 이용한 인식실험을 추가적으로 실시하였다.

표 2에서 보듯이 플랫렉시콘에서는 ECHOS의 인식률이 HTK와 거의 비슷한 결과가 나왔고, 인식속도는 ECHOS가 HTK보다 2.5배 정도 느렸다. ECHOS가 객체 지향적 구조와 STL을 이용한 C++ 언어로 되어 있고, 프로그램 구현상의 차이로 인하여 인식속도의 차이가 발생한다. ECHOS에서 지원되는 렉시컬트리 탐색에서는 ECHOS가 HTK보다 0.5배 정도 빠르게 인식하지만, 40%정도의 상대적 에러율 증가를 나타내고 있다. 그 이유는 ECHOS의 렉시컬트리 탐색에서는 그림 7과 같이 하나의 렉시컬트리를 사용하고, 단어간의 천이가 일어나는 순간에 언어모델이 적용되는 플랫렉시콘 탐색과는 달리 렉시컬트리 탐색에서는 인식될 단어가 결정되는 단어 끝부분에서 언어모델이 적용되어, 단어가 시작되는 노드에 잘못된 누적확률이 적용되거나 인식되어야 될 단어가 먼저 가지치기 (pruning)되는 경우가 발생하기 때문이다.

렉시컬트리에서의 에러율 증가를 막기 위하여 바이그램 서브트리를 복사하는 방법이 제안되어 있으나, 메모

리 크기 및 프로그램 복잡도 측면에서 렉시컬트리의 이점이 상쇄된다[13]. 계산량 및 메모리 증가를 억제하면서도 인식률을 개선하는 다른 방법으로서, 렉시컬트리의 잎사귀노드 (leaf node)로 진입할 때 언어모델을 적용하거나, 단일 음소 단어들을 플랫폼의 형태로 특별히 처리하는 방법[14]이 있다. 이것은 언어모델을 가능하면 조기에 적용함으로써 음향모델 스코어가 낮은 정확한 가설 (correct hypothesis)이 가지치기되는 것을 막을 수 있기 때문이다. 이 방법들은 ECHOS 차기 버전에서 보완될 예정이다.

앞에서 제시한 비교실험에서는 단어내 (intraword)모델만을 적용하였다. 이는 ECHOS 현재 버전이 단어간 (interword)모델을 지원하지 않기 때문에 동일 조건의 비교실험이 불가능하기 때문이다. 실제로 HTK에 단어간 모델링을 적용할 경우 인식률은 더 높게 나타난다.

V. 결론

교육 및 연구를 위하여 개발된 한국어 음성인식 플랫폼의 구조, 기능, 구현 및 검증 결과를 소개하였다. 개발된 플랫폼은 쉽고 작으면서 한글 처리가 가능한 객체기반의 구조를 가진다. 소규모 FSN 및 8000단어 연속음성 인식 태스크에 대하여 공개 음성인식기인 HTK와 성능을 비교하였다. FSN에서는 HTK와 유사한 인식률을 나타내었으며, 객체지향 구조에 의하여 실행시간이 2배로 증가하였다. 그러나 8000단어 연속음성인식에서 ECHOS는 렉시컬트리 탐색 알고리즘을 채용함으로써 에러율은 증가하지만 인식시간을 HTK의 0.5배로 감소시켰다. 차기버전에서는 렉시컬트리에서의 단어간모델링, 단일음소 단어 처리, 화자적응, 클래스 n-그램 기능을 추가하여 인식률을 향상할 계획이다. 이 플랫폼은 국내 음성인

식 분야의 저변을 확대하고, 연구자들이 알고리즘 연구에 전념할 수 있는 토대를 마련하며, 음성인식기술의 비교 기준의 역할을 할 것으로 기대된다.

감사의 글

이 논문은 음성정보기술산업지원센터의 연구비 지원으로 휴먼인터페이스연구조합을 통하여 “한국어 음성인식 플랫폼 개발” 과제에서 수행한 내용입니다.

참고 문헌

1. HTK Home page. <http://htk.eng.cam.ac.uk>
2. CMU Sphinx: Open Source Speech Recognition. <http://www.speech.cs.cmu.edu/sphinx/Sphinx.html>
3. Automatic Speech Recognition: Software. <http://www.isip.msstate.edu/projects/speech/software/>
4. Multipurpose Large Vocabulary Continuous Speech Recognition Engine Julius. <http://www.ar.media.kyoto-u.ac.jp/members/ian/doc>
5. <http://speech.chungbuk.ac.kr/~owkwon/srhome/index.html> ezCSR.
6. 권오욱, 김희린, 유창동, 김봉완, 이웅주. “한국어 음성인식 플랫폼의 설계.” *말소리*, 51 (9), 2004.
7. Standard Template Library Programmer’s Guide. <http://www.sgi.com/tech/stl/>
8. Practical UML: A Hands-On Introduction for Developers- by Randy Miller.
9. L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*, (Prentice-Hall, 1993).
10. F. Jelinek, *Statistical Methods for Speech Recognition (Language, Speech, and Communication)*, (MIT Press, 1999).
11. S.B. Davis and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,” *IEEE Trans. ASSP*, 28, 357-366, Aug. 1980.
12. H. Hermansky, “Perceptual linear predictive (PLP) analysis of speech,” *Journal of the Acoustical Society of America*, 87 (4), 1738-1752, 1990.
13. Aurora, Distributed Speech Recognition. <http://portal.etsi.org/sto/kta/DSR/dsr.asp>
14. X. Huang, A. Acero, and H.-W. Hon, *Spoken Language Processing*, 648-650, Prentice Hall, 2001.
15. M.K. Raishankar, *Efficient Algorithms for Speech Recognition*, (PhD Thesis, CMU, 1996).

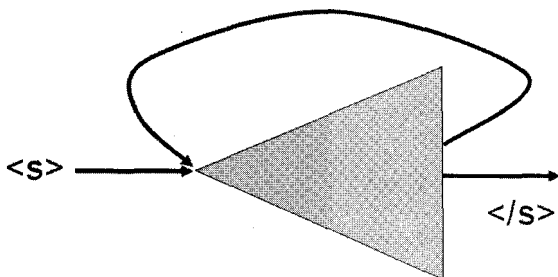


그림 7. ECHOS에서의 렉시컬트리 구조
Fig. 7. A lexical tree for large vocabulary continuous speech recognition in ECHOS.

저자 약력

• 권오욱 (Oh-Wook Kwon)



1986년 2월: 서울대학교 전자공학과 졸업
 1988년 2월: 한국과학기술원 전기및전자공학과 (공학석사)
 1997년 2월: 한국과학기술원 전기및전자공학과 (공학박사)
 1986년~2000년: 한국전자통신연구원 책임연구원
 1999년 4월~1995년 12월: 독일 Karlsruhe University 방문연구원

2000년~2001년: 한국과학기술원 연구교수
 2001년~2003년: Postgraduate Research Associate, University of California, San Diego
 2003년 9월~현재: 충북대학교 전기전자컴퓨터공학부 조교수

• 권석봉 (Sukbong Kwon)



1997년 2월: 연세대학교 수학과 졸업
 1999년 2월: 한국과학기술원 수학과 (이학석사)
 1999년~2002년: PAXVR 음성탐 연구원
 2004년 8월: 한국정보통신대학교 공학부 (공학석사)
 2004년 8월~현재: 한국정보통신대학교 공학부 박사과정

• 장규철 (Gyuchoel Jang)

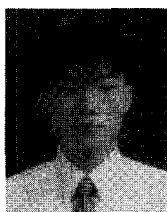
2001년 2월: 한국과학기술원 전기및전자공학과 졸업
 2003년 8월: 한국과학기술원 전기및전자공학과 졸업 (공학석사)
 2003년 8월~현재: 한국과학기술원 전기및전자공학과 (박사과정)

• 윤성락 (Sungrack Yun)



2003년 2월: 한국과학기술원 전기및전자공학과 졸업
 2003년 3월~현재: 한국과학기술원 전기및전자공학과 석·박사통합과정

• 김용래 (Yong-Rae Kim)



2003년 2월: 우송공업대학 전자정보계열 컴퓨터응용과 졸업
 2005년 2월: 충북대학교 컴퓨터공학과 졸업
 2005년 3월~현재: 충북대학교 제어계측공학과 석사과정

• 장광동 (Kwang-Dong Jang)



2001년 2월: 공주대학교 전기전자정보통신공학부 전기공학전공 졸업
 2001년~2004년 8월: 다원비전
 2004년 9월~현재: 충북대학교 제어계측공학과 석사과정

• 김희린 (Hoi-Rin Kim)



1984년 2월: 한양대학교 전자공학과 졸업
 1987년 2월: 한국과학기술원 전기및전자공학과 (공학석사)
 1992년 2월: 한국과학기술원 전기및전자공학과 (공학박사)
 1987년 10월~1999년 12월: 한국전자통신연구원 선임연구원

1994년 6월~1995년 5월: 일본 ATR-ITL 방문연구원
 2001년 1월~현재: 한국정보통신대학교 공학부 부교수

• 유창동 (Changdong Yoo)

제20권 3호 참조

• 김봉완 (Bong-Wan Kim)



1995년 2월: 원광대학교 컴퓨터공학과 (공학사)
 1997년 2월: 원광대학교 컴퓨터공학과 (공학석사)
 2002년 2월: 원광대학교 컴퓨터공학과 (공학박사)
 2001년 1월~현재: 음성정보기술산업지원센터 연구기획실장

• 이용주 (Yong-Ju Lee)



1976년 2월: 고려대학교 전자공학과 (공학사)
 1987년 2월: 고려대학교 전자공학과 (공학석사)
 1992년 2월: 고려대학교 전자공학과 (공학박사)
 1976년~1980년: 공군 제7항로보안단 통신전자장교
 1980년~1994년: 한국전자통신연구소 지동통역연구실 실장 (책임연구원)

1994년~현재: 원광대학교 전기 전자 및 정보공학부 교수
 2001년~현재: 음성정보기술산업지원센터 센터장