

음성인식 및 합성을 이용한 십자말 게임의 구현

김동주, 윤재선, 이용주, 김동환, 홍광석
성균관대학교 전기전자컴퓨터공학부 휴먼컴퓨터연구실

An Implementation of Crossword Game using Speech Recognition and Synthesis System

Dong-Ju Kim, Jeh-Seon Youn, Yong-Ju Lee, Dong-Hwan Kim, Kwang-Seok Hong
HCI Lab, Electrical & Computer Engineering, Sungkyunkwan University

radioman@hci.skku.ac.kr, sunhci@ece.skku.ac.kr, blueyong@popmail.com, kdh2k@popmail.com, kshong@yurim.skku.ac.kr

요약

본 논문에서는 연구실에서 만든 음성인식기와 합성기를 이용하여 십자말 게임을 구현하였다. 십자말 게임에는 고사성어 600개 정도의 단어가 사용되었으며, 다른 영역별 사전을 추가 할수 있도록 만들어졌다. 구현된 게임은 시작, 진행 등의 모든 과정이 음성으로 동작하며, 부파적인 정보는 음성 합성(TTS)에 의해 이루어진다.

십자말 게임에 사용되는 단어의 배열은 매번 랜덤하게 선택되도록 구성되며, 음성 인식기는 VCCV (Vowel + Consonant + Consonant + Vowel) 기반의 화자독립으로 구현되었다. 선택된 문제에 대한 설명은 텍스트로 보여주면서, 동시에 TTS 시스템에 의해 음성으로 출력된다.

I. 서론

사용자와 컴퓨터의 보다 편리한 인터페이스를 위하여 음성인식 및 합성을 이용한 사용자 인터페이스는 최근 널리 연구되고 있다. 컴퓨터와 인간의 인터페이스 기술로 키보드나 마우스가 많이 사용되고 있지만, 보다 자연스럽고 편리한 음성을 이용한 인터페이스 기술이 적극적으로 개발되고 활용되고 있다. 음성을 이용한 인터페이스 기술은 여러 분야에 적용되어 서비스되고 있지만 응용분야의 부족 및 많은 개선점이 필요하다.

따라서 본 논문에서는 일반인들이 널리 즐길 수 있는 게임을 선택하여 음성인식의 응용 측면을 좀 더 부각시키고자 한다. 본 논문에서 구현된 십자말 게임에는 음성인식과 합성기술이 적용되어 더욱더 편리한 인터페이스를 제공한다. 십자말 게임은 현재 고사성어 단어들에 대하여 구성되었으며 다른 분야의 단어들도 적용 가능하도록 설계되었다. 음성인식 및 합성을 적용한 십자말 게임은 게임의 측면뿐만 아니라 영역별 사전만을 추가 함으로써 쉽고 다양한 학습 효과도 기대할 수 있다.

II. 십자말 게임의 구성

십자말 게임은 펜티엄 PC이상, 사운드카드, 윈도우 2000의 PC환경에서 구현되었다. PC에서의 사용자 인터페이스는 Visual C++의 MFC를 이용하여 구현되었다. 게임은 미리 준비된 영역별 사전의 단어 배열로부터 시작된다. 단어의 배열은 매번 랜덤하게 생성되며 창틀의 크기를 선택하여 단어 배열의 크기를 조절할 수 있도록 구성되었다. 십자말 배열의 구성이 이루어지면 사용자는 마우스나 음성을 통하여 문제를 선택한다. 문제를 선택하면 선택된 문제에 대한 설명이 윈도우의 디스플레이 장치에 표시되며, 동시에 TTS시스템에 의해 설명이 음성으로 출력된다. 문제의 설명을 보고 사용자는 문제의 정답을 키보드로 입력하거나 음성을 통해 맞출 수 있도록 게임은 구성되었다.

PC에서 구현된 십자말 게임의 전체적인 구조는 그림 1과 같다.

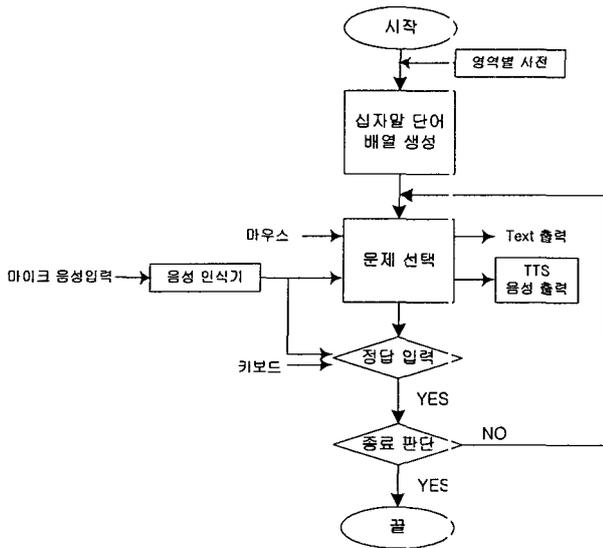


그림 1. 십자말 게임의 구성

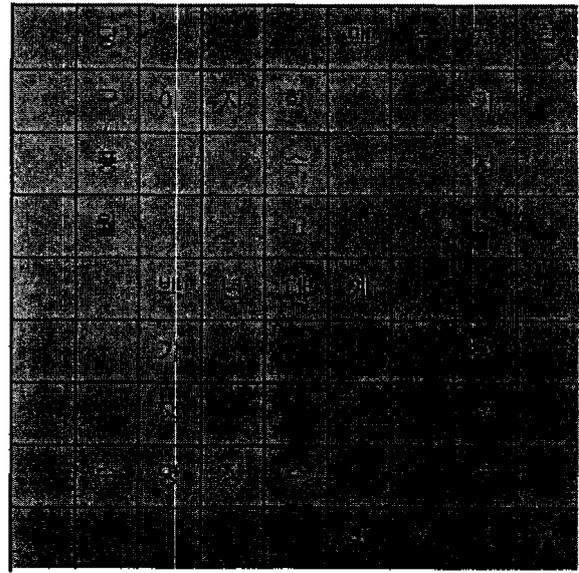


그림 2. 교사성어 낱말의 배열

2.1 십자말 단어의 배열

십자말 게임에 사용되는 단어는 주어진 영역별 사전을 이용하여 선택된 창틀에 맞도록 랜덤하게 배열된다. 처음 선택되는 단어는 랜덤하게 선택되나 배치되는 위치도 랜덤하게 선택된다. 이것은 매번 게임을 할 때마다 서로 다른 십자말 배열을 갖게 한다. 처음 선택된 단어를 기준으로 단어의 배열은 창틀내에 공간이 부족할 때까지 연결단어를 찾게 된다. 연결단어가 부족할 때에는 새로운 빈공간을 탐색하여 다시 그 위치로부터 연결 단어를 생성하게 된다. 연결 단어의 배열을 구성하는데 탐색 시간이 필요하지만 매 시간마다 다른 연결의 배열을 생성하는 장점을 갖는다.

단어 배열의 크기는 사용자가 임의로 선택할 수 있도록 구성되었다. 프로그램에서는 9×9, 12×12, 15×15, 18×18 크기의 창틀을 제공한다. 창틀의 크기가 커질수록 연결되는 단어의 개수와 탐색 시간은 증가하게 된다.

본 프로그램에서는 교사성어 600여개의 레이터를 사용하여 십자말 게임을 구성하였다. 교사성어의 데이터뿐만 아니라 다른 영역의 사전만 추가하면 새로운 영역의 십자말 게임이 가능하도록 프로그래밍 되었다.

그림 2는 교사성어 사전을 이용한 9×9 창틀에서의 십자말 배열을 보여준다.

2.2 음성인식 시스템

낱말의 배열이 이루어지고 게임이 시작되면 동작의 모든 부분은 키보드나 마우스뿐만 아니라 음성으로 제어할 수 있게 된다. 음성에 의한 동작은 메뉴제어, 문제선택, 정답선택 등의 제어에 이용된다. 또한 상황에 따라 인식되는 명령어는 다르게 동작한다. 예를 들어 배열이 아직 구성되지 않는 초기에는 메뉴의 명령어만 인식되며, 배열이 구성이 완료되면 메뉴의 명령어와 문제 번호, 배열에 사용된 교사성어만을 제한하여 인식하게 된다.

마이크로부터 입력되는 신호는 음성신호뿐만 아니라 잡음신호도 같이 들어오게 된다. 잡음으로 인한 오인식을 피하기 위하여 음성 부분만을 검출하는 것이 필요하다. 음성의 시작점 및 끝점 검출은 연속적인 입력신호의 에너지값과 설정된 임계값의 비교를 통하여 이루어진다.

본 논문에서 사용된 음성 인식단위는 비교적 분할작업을 수월하게 할 수 있고, 인식시 분할작업에서 발생하는 약간의 오류를 수용할 수 있는 CV, VCCV, VC를 사용하였다. 음성 데이터 베이스로는 본 연구실에서 구축한 118개의 성이 포함된 성명데이터 1145개, 단음절 521개, PBW데이터가 포함된 1001개, FDB 4088개를 사용하였다. 훈련데이터로부터 CV단위 308개, VCCV단위 3319개, VC단위 56개를 추출하여 Reference Model을 구성하였다.[1] 인식 단어의 모델을 생성할 때 VCCV모델로 구성되지 않는 모델은 VC와 CV모델을 연결하여

구성하도록 시스템은 설계되었다.

인식단계에서는 검출된 음성으로부터 Mel-Cepstrum 특징벡터를 추출하고, VQ코드북을 통과하여 인덱스 열을 얻는다. 이로부터 미리 만들어진 영역별 사전의 단어에 대한 Reference Model과 확률값을 각각 비교하여 인식 단어를 결정한다.[2] 인식과정은 그림 3과 같다.

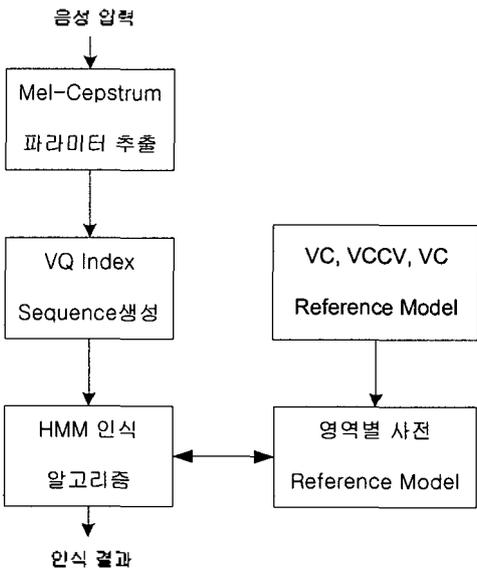


그림 3. 음성인식 시스템의 구성

프로그램의 명령어 메뉴에는 “게임, 도움말”이 있으며, “게임” 명령어의 서브메뉴로는 “시작, 맵선택, 종료”,가 있고 “도움말” 명령어의 서브메뉴로는 “정보, 만든태사람들”이 각각 있다. 게임에서의 음성인식은 각 상태에 따라 인식 후보를 다르게 하여 인식률을 높게 하였다. 즉, 프로그램 시작시에는 명령어 “게임”과 “도움말”만이 인식이 되며, “게임”이 인식되면 인식후보의 단어는 “시작, 맵선택, 종료”로 제한된다.

인식 대상의 단어는 메뉴의 명령어뿐만 아니라 문제 선택시 “일번, 이번...”과 같은 문제 순서 번호와 문제의 정답이 되는 단어들이 있다. 게임에 필요한 십자말 단어 배열이 완성되면 음성인식기는 위의 모든 단어를 인식 대상으로 갖는다.

본 논문에서 영역별 사전은 고사성어 단어들로 구성된다. 고사성어 데이터로부터 구성된 Reference Model은 인식시 모든 후보를 인식의 대상으로 하지 않는다. 즉, 인식 시간의 단축 및 인식률의 향상을 위하여 게임 배열에 사용된 단어만을 인식의 단위로 제한하게 된다.

2.3 음성합성 시스템

무제한 음성합성 시스템은 어떠한 문장이라도 음성으로 변환하여 출력해 내는 시스템(TTS)을 말한다. TTS 시스템은 크게 데이터 베이스를 구축하는 부분과 실제로 음성합성을 수행하는 부분으로 나눌 수 있다. 음성 데이터 베이스는 남녀 화자 각각 한 명씩 빈도수가 높은 어절 1000 여개와 반응절 단어 452개를 발생시켜 구축하였다. 합성 방식은 TD-PSOLA 방법을 적용시켰으며, 합성 단위는 반응절과 어절을 사용하였다. 합성 과정은 그림 4와 같다.

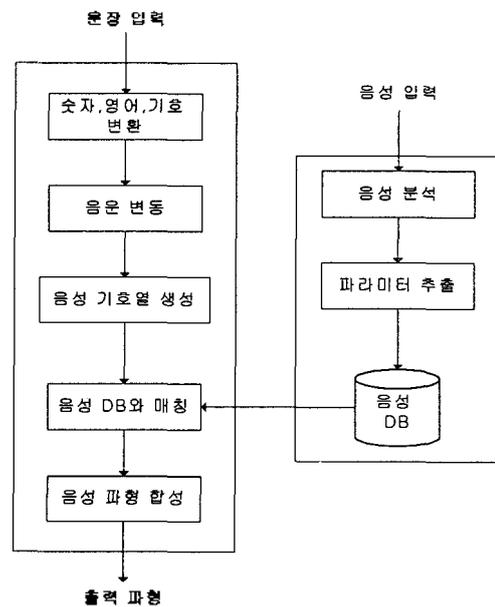


그림 4. 음성 합성 시스템의 구성

우선 문장이 입력되면 숫자나 영어에 대한 변환과정을 거친 후, 한국어로 소리나는 대로 변환시키는 음운 변동 과정을 거친다. 이로부터 음성 기호열을 생성한 후 사전에 구축되어진 음성 DB로부터 해당하는 소리를 가져와 음성 합성을 하게 된다.[3] 프로그램에서 음성합성 시스템은 선택된 문제 단어에 대한 설명과 정답을 맞추었는지 여부를 들려주는데 이용된다.

2.4 십자말 게임의 실행

음성인식과 합성이 결합된 십자말 게임 프로그램의 실행화면은 그림 5와 같다.

단어의 배열은 각각 번호가 붙여져 있으며, 음성이나 마우스를 이용하여 문제를 선택할 수 있다. 문제를 선

택하면 선택된 문제에 대한 설명이 텍스트로 주어지며 TTS에 의해 음성으로 출력된다.

선택된 문제의 설명을 보고 키보드나 음성으로 정답을 맞추게 되면 TTS시스템에 의해 “정답입니다” 라는 음성이 출력되며, 틀린 경우에는 “틀렸습니다” 라는 음성이 출력되어 정답을 맞추었는지 여부를 알 수 있도록 하였다. 정답을 맞추었을 경우에는 정답이 텍스트로 표시되어 더 이상 선택을 할 수 없게 된다.

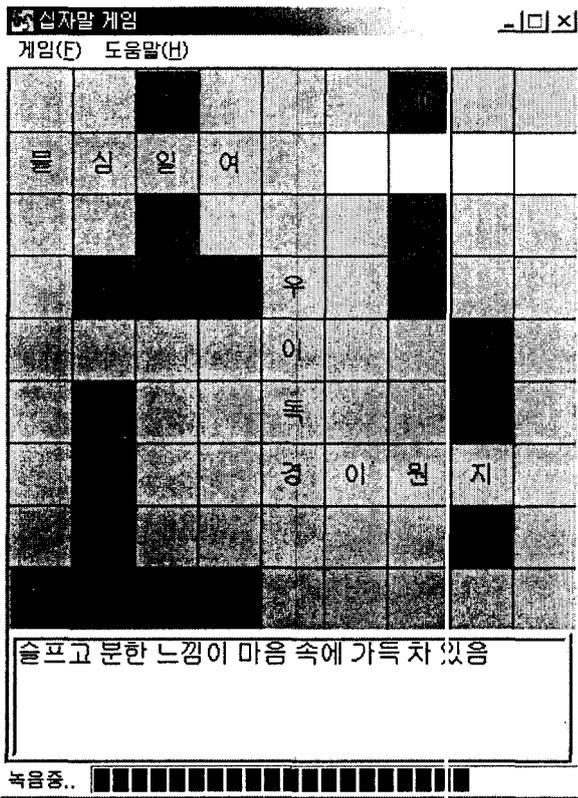


그림 5. 십자말 프로그램의 실행

III. 인식실험 및 결과

인식 실험에는 총 5명의 남성화자가 참여하여, 각각의 창틀에 따라 생성된 단어 배열을 인식하는 실험을 하였다. 이때 인식에 사용되는 배열은 각 화자마다 다르게 구성하였다.

표 1은 십자말 게임의 창틀에 따른 인식율 측정 결과를 보여준다. 각 창틀에 따라 5명의 화자가 인식실험을 하였으며 최종 인식율을 보여준다. 실험 결과로부터 창틀의 크기가 커질수록 인식 비교대상이 많아지므로 인식율이 약간 저하됨을 알 수 있다.

표 1. 십자말 게임의 인식 실험

	9×9	12×12	15×15	18×18
화자 1	9/9	14/14	22/25	30/35
화자 2	12/12	14/14	25/26	33/34
화자 3	9/9	16/17	29/29	32/33
화자 4	11/11	17/17	27/27	31/32
화자 5	11/11	14/14	27/28	31/32
평균인식률	100(%)	98.6(%)	96.2(%)	94.5(%)

IV. 결론

본 논문에서는 음성인식과 합성을 이용한 십자말 게임의 구현에 대해 간략히 다루어보았다. 음성을 이용한 십자말 게임은 일반 상식 및 다양한 정보를 게임을 통하여 좀더 쉽게 습득할 수 있으리라 사료되며, 음성을 이용한 게임 및 어학학습 등의 활용분야에 크게 기여할 것으로 생각된다.

음성인식 프로그램 제작에 있어 가장 큰 문제로 대두되는 인식률 문제는 기존에 구축해 놓은 데이터 베이스의 양이 음성인식을 하는데 있어 충분히 컸기 때문에 인식면에서는 크게 문제가 되지 않았다. 하지만 십자말 단어 배열의 검색 시간을 단축하는 새로운 알고리즘과 언어 필요하다.

참고문헌

- [1] 윤재선, 정광우, 홍광석, “무제한 단어인식 시스템을 위한 VCCV분할에 관한 연구,” 한국음향학회 하계 학술발표대회 논문집, 제 19권 1호, 2000.
- [2] 윤재선, 홍광석, “무제한 어휘 독립 단어 인식 시스템의 구현”, 음성통신 및 신호처리 학술대회 논문집 제 17권 제1호, pp. 145-148, 2000.
- [3] 양원렬, 윤재선, 홍광석, “영한 음차 변환을 이용한 무제한 음성인식 및 합성기의 구현”, 한국 신호처리 시스템학회 논문집 제1권 제1호, pp181-184, 2000.