

# 음성 조합 및 인식 기술을 이용한 영어 학습 프로그램

강지훈, 이종범, 강경구, 한재일  
국민대학교 컴퓨터공학부

## English Learning Program Based on Speech Synthesis and Recognition

Han, Jae-Il, Lee, Jong-Bum, Kang, Kyung-Koo, Kang, Ji-Hoon  
School of Computer Science, Kookmin University  
E-mail : hicracks@gmail.com

### 요 약

세계화 시대를 맞아 영어 교육에 대한 필요성은 점차 증가하고 있다. 이에 따라 아동기 영어 교육에 대한 관심이 점차 증가함과 동시에 잘못된 교육 방식에 의한 부작용 또한 증가하고 있다. 아동기의 영어 학습 프로그램은 공부가 아닌 놀이의 측면으로 접근하는 것이 중요하다. 본 논문은 기존의 주입식 영어 학습을 탈피하고 놀이로써 접근하기 위한 음성 조합 및 인식을 이용한 양방향 영어 학습 프로그램의 설계 및 구현에 대하여 논한다.

### 1. 서론

영어는 미국, 영국, 호주, 뉴질랜드, 캐나다 등의 수많은 곳에서 모국어로 사용되고 있다. 그리고 유럽 연합의 공용어이자, 홍콩, 동남아시아 인도네시아 일부 지역, 일부 중남미의 국가에서 공용어로서의 역할을 담당해내고 있다. 또한 대부분의 지구촌 국가에서 영어를 공부하고 있다. 영어는 세계화 시대의 필수 경쟁력인 것이다.

대다수의 부모들은 이런 이유들 때문에 영어 조기교육을 중요시하고 이에 따라 영어교육 시장은 날로 커져 10조원 대에 육박하고 있다. 하지만 영어 교육 시장이 증대될수록 아이들의 영어 거부증이라는 단점 또한 부각되고 있다.

영어 거부증의 대표적인 원인은 영어 학습 방법에 있다. 워크북 중심의 학습 형태, 숙제가 많은 학원에 다니는 아이들이 영어 거부증을 더 많이 보인다는 연구가 있으며[1], 녹음된 음성이나 동영상의 재생을 통한 반복 및 주입식 교육이 원인임

을 보이고 있다. 따라서 아동기의 영어 학습 프로그램은 공부가 아닌 놀이로 다가가는 것이 중요하다.

이에 본 연구는 아동이 친근감을 느끼는 GUI 인터페이스를 디자인하고, 음성 조합 및 인식 기술을 이용하여 듣기와 말하기 연습뿐 아니라 임의의 문장에 대한 실시간 듣기와 말하기 연습을 제공하는 영어학습기를 설계하고 구현했다.

본 논문의 구성은 다음과 같다. 2장은 기존에 연구되어 왔던 음성인식의 기술 동향 및 아동 영어 학습기 관련 연구 등을 설명하며, 3장에서는 음성 조합 및 인식을 이용한 양방향 영어 학습 프로그램의 설계 및 구현에 대하여 기술한다. 4장은 결론 및 앞으로의 연구 방향을 논한다.

### 2. 관련연구

음성인식은 마이크나 전화 등을 통해 얻어진 음향학적 신호를 단어나 단어 집합 또는 문장으로 변

환하는 과정을 말한다.[2] 음성인식에 관한 연구는 미국, 유럽, 일본 등 선진국뿐만 아니라 국내에서도 활발히 진행 중이다.

미국에서는 1970년 국방성의 주도로 시작하여 현재 주요 연구기관에서 잡음환경에 강인한 음성인식기술, 모바일에 적합한 음성인식 시스템 개발, 대화체 음성인식을 위한 인식 기법 및 모델의 개발이 진행되고 있고 유럽에서는 유럽연합의 주도로 언어 공학 및 인간 언어기술 과제를 수행하고 있다. 일본에서도 회자 독립 음성인식, 자유 발화 음성인식 등에 대한 연구가 진행 중이다.[3]

국내의 대표적인 음성인식 기술 동향은 한국전자통신연구원(ETRI)에서 HCI를 위한 음성 입/출력 처리에 관한 연구와 한국통신에서 증권정보 안내시스템이 있다[3].

영어 학습 프로그램에 관한 관련 연구는 다음과 같다.

#### 1) 멀티미디어 전자영어학습 교재

방과 후 학습에 사용하기 위해 개발한 교재로써 수준별 학습, 자기주도적 학습, 상호작용을 통한 언어 능력 향상 등을 목표로 한다. 레슨형, 스토리텔링의 두 형식으로 100시간 단위로 언어의 4기능을 차시별로 구분하여 듣기, 말하기, 읽기, 쓰기, 심화, 보충의 5차시로 이루어져 있으며 각 차시는 각 영역과 관련된 4-5가지 학습활동으로 구성된다.

서책형 교과서의 기본 기능을 재현하기 위해 서책형 이미지 메타포를 활용했다. 이를 통해 전체적인 내용과악이 쉽고 수업과 보다 긴밀히 연계되는 장점을 갖는다. 또한 동영상 세대에 우려될 수 있는 글 읽는 능력의 감소를 막는다.[4]

### 3. 본론

본 프로그램의 대상 연령은 미취학 아동부터 초등학교 저학년이므로 인터페이스가 간단하고 컨트롤이 쉬워야 한다. 또한 사용자로 하여금 흥미를 유발해서 공부가 아닌 놀이로 인식되어야 한다. 최대한의 학습 효과를 이끌어내기 위해 균형 잡힌 교육을 제공해야 한다. 이를 위해 다음과 같은 요구사항이 필요하다.

#### 3.1 요구사항

첫 번째로 아동들에게 친근감을 주고 쉽게 사용

할 수 있는 GUI 인터페이스가 요구된다. 두 번째로 말하기, 듣기, 쓰기 등 균형잡힌 교육 콘텐츠를 제공해야 한다. 세 번째로 음성 조합 및 인식을 사용하여 사용자의 흥미를 유발하고 효율적인 학습을 제공할 수 있어야 한다.

##### 3.1.1 유즈케이스

본 프로그램에서는 Listening, Speaking, Dictation, Q&A의 네 가지 콘텐츠를 제공한다. 그림 1은 이것을 표현한 유즈케이스 다이어그램이다.

Listening 부분은 미리 입력된 텍스트를 읽어주는 것이 주된 기능이다. 이를 통해 사용자는 청각적 학습이 가능하다. 그림 1에서 전체내용듣기에 해당하는 것으로 부분 또는 전체 문장에 대한 반복 청취가 가능하여 잘못들은 문장에 대한 반복학습 효과를 준다. Speaking 부분은 다이어그램의 Role play에 해당하는 것으로 챗터에 대한 역할을 선택하여 사용자가 해당 역할에 대한 대사를 읽는 부분이다. 사용자와 외국인이 가상으로 대화하는 상황을 연출하여 외국인과의 대화에 대한 두려움을 극복할 수 있도록 한다. Dictation 부분은 Listening만으로는 부족할 수 있는 부분을 보완해 준다. 단순히 듣는 것을 넘어서 정확한 스펠링을 알고 있는지 또한 제대로 Listening을 하고 있는지 확인할 수 있는 학습내용을 제공한다. Q&A는 실생활에서 사용하는 대화형식을 문제로 풀어보는 부분이다. 문제에 대한 보기 중에서 답을 선택해 말하면 정답 유무를 알린다. Listening, Speaking, Dictation, Q&A의 네 가지 콘텐츠를 제공함으로써 말하기, 듣기, 쓰기 등의 균형 잡힌 교육이 가능하도록 설계했다.

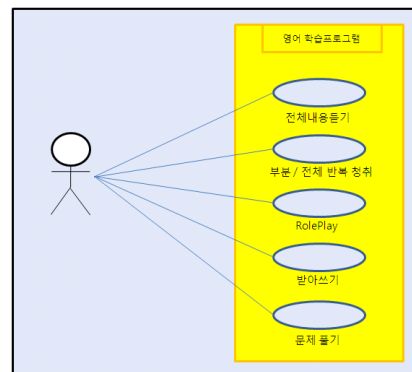


그림 1. 유즈케이스 다이어그램

### 3.1.2 GUI 인터페이스

GUI 인터페이스는 Visual studio C#과 photoshop을 사용했다. 아동에게 친숙하게 보이기 위해 아동 취향의 색을 고려하여 구성한다. 초등학교 저학년 대상의 색채 선호도 중 상위 50%에 해당하는 색채를 위주로 디자인함으로써 프로그램에 친숙함을 더하도록 했다.[5] 모든 기능은 사용자가 쉽게 컨트롤 할 수 있도록 구조를 단순화했다.

### 3.1.3 음성 조합 및 인식

사용자의 흥미를 유발하여 효율적인 교육을 제공하기 위해 MS에서 제공하는 SAPI를 사용했다. 음성인식 기능을 사용하여 대화 및 문제 풀이 기능을 사용한다. 그림 2는 음성인식 시스템의 기본 구성도를 나타낸다. 먼저 사용자가 마이크를 통해 해당 문장에 대한 음성을 입력하면 음성특징분석을 통해 음성특징계수들로 추출된다. 이것들은 패턴인식과정으로 넘겨져 해당 문장과 비교되며 해당 문장과 동일할 경우 음성을 인식함으로써 진행된다[6].

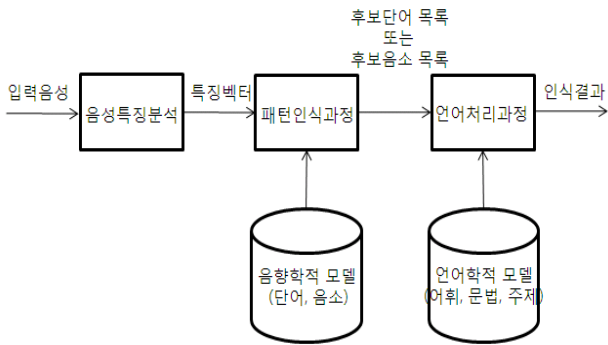


그림 2. 음성인식 시스템 기본 구성도

### 3.2 설계 및 구현

그림 3은 프로그램에 대한 메인화면에 대한 시퀀스 다이어그램이다. 사용자의 요청이 들어왔을 때 실제 내부에서 어떻게 상호작용하는지를 보여준다.

다음은 Listening 모드를 예로 설명한 시스템의 내부적인 동작이다. 그림 4는 Listening 모드에 대한 시스템 시퀀스 다이어그램이다. Listening 모드는 다음과 같이 작동한다. 우선 사용자는 프로그램

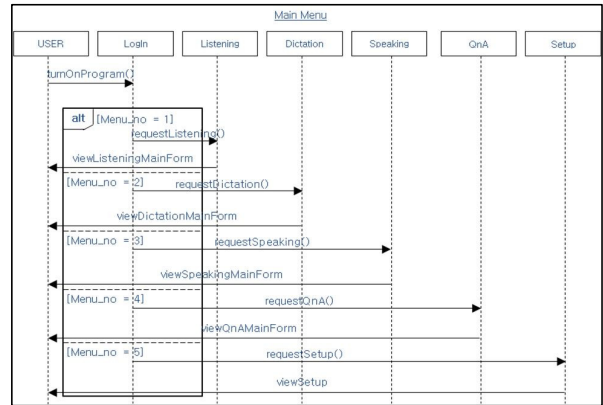


그림 3. 메인 화면에 대한 시퀀스 다이어그램

을 실행시킨다. 네 가지 모드 중 Listening 모드를 선택한다. 시스템은 Listening 모드에 해당하는 메뉴를 사용자에게 보여준다. 사용자는 원하는 챕터를 선택한다. 시스템은 해당 챕터에 대한 텍스트를 화면에 출력한다. 사용자는 컨트롤 버튼 중 재생 버튼을 선택하면 시스템은 문장을 재생하고 재생하는 문장에 해당하는 그림을 메인 화면에 출력하는 구조로 Listening 모드가 진행된다. 다른 세 콘텐츠들도 이와 같은 구조로 이루어져 있다.

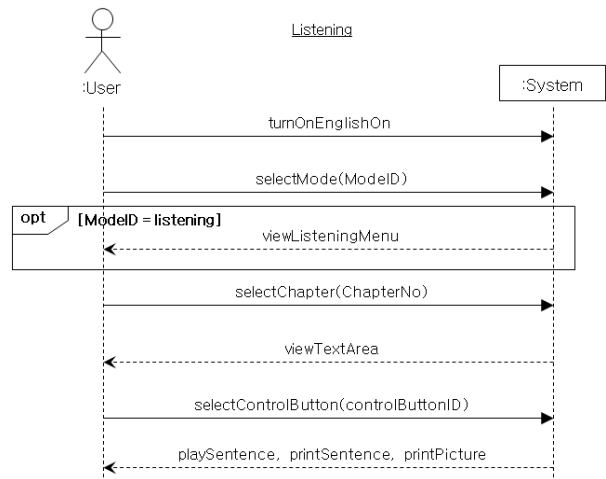


그림 4. Listening 모드에 대한 시스템 시퀀스 다이어그램

### 3.3 실행 화면

그림 5는 영어 학습 프로그램의 데모 화면이다. 작동 방법은 다음과 같다. 네 가지 콘텐츠 중 하

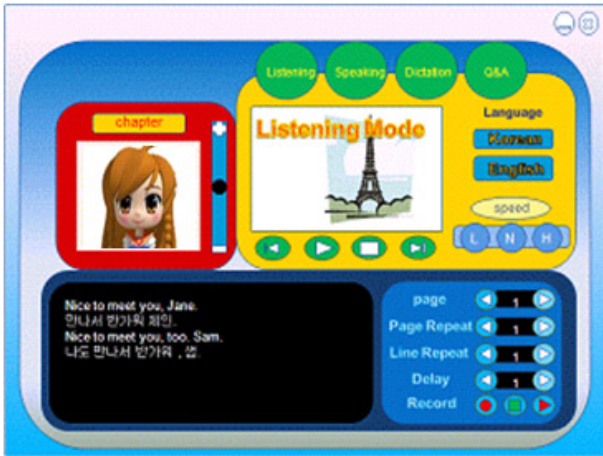


그림 5. 영어 학습 프로그램 GUI 인터페이스

나를 선택한다. 화면 상단의 챕터 버튼을 선택하여 원하는 챕터를 선택한 뒤 play 버튼을 누르면 해당 콘텐츠에 대한 내용이 실행된다.

모든 콘텐츠의 챕터에 대한 대사와 관련된 화면을 출력함으로써 영, 유아기 아동들의 호기심과 흥미를 유발한다. 영어 발음과 입술 그림을 동기화하여 영어 단어 발음에 대한 입술 모양을 출력함으로써 청각적 효과와 시각적 효과를 동시에 얻을 수 있다. 뿐만 아니라 영어에 익숙하지 않은 아동들을 고려해 재생되는 음성의 속도를 3단계로 나누어 설정할 수 있도록 했다. 음성인식을 통한 양방향 학습은 아이들의 적극적인 참여를 도와 학습 능력 증진은 물론 영어에 대한 거부감도 함께 없앨 수 있다. 음성인식을 사용함으로써 사용하지 않았을 때와 달리 자신의 스피킹 수준을 확인함과 동시에 발음 연습이 가능하다. 또한 대화 기능을 통해 실용성을 강조했고 효과적인 유아 영어 교육을 제공한다.

#### 4. 결론 및 향후 연구 방향

영어 교육의 궁극적인 목표는 외국인과의 원활한 의사소통이다. 이것을 위해 많은 인적, 물적 자원을 투자하고 있지만 많은 비용에 비해 적은 효율에 대한 논란이 나타나고 있다. 많은 시간과 돈을 투자하여 어릴 때부터 교육을 하더라도 이것에 대한 흥미를 느끼지 못한다면 아무 소용이 없기 때문이다. 본 논문에서는 사용자가 영어를 쉽고 친근하게 느껴 저비용의 투자로 고효율을 얻기 위해

음성인식을 통한 양방향 영어 학습 프로그램을 제안했다.

본 논문에서 제안한 프로그램은 사용자가 원하는 시나리오를 예측하여 구현했다. 또한 사용자가 쉽게 사용할 수 있도록 사용자의 입장에서 GUI 인터페이스를 구성했다.

하지만 주변의 소음으로 인한 인식률 저하와 사용자의 다른 인식률의 차이가 크다는 문제점이 있다. 이런 문제점을 해결하기 위해 인식률 개선에 대한 지속적인 연구가 필요하다.

#### [참고문헌]

- [1] 한국교육평가원, <http://e.kice.re.kr/>
- [2] Paul Dalsgaard, Zheng-Hua Tan and Borge Lindberg, Partial Splicing Packet Loss Concealment for Distributed Speech Recognition, 2003
- [3] Team Delta, Technical Strategy for Mobile Speech Recognition, 2005
- [4] 진경애 외 10명, 멀티미디어 전자영어학습 교재 개발 사업 완료 보고서, 한국교육과정 평가원, 2007
- [5] 최정화, 아동의 색채선호도와 색채지도에 관한 연구, 단국대학교 교육대학원, 2004
- [6] 음성인식(Automatic Speech Recognition)기술과 특허동향, 2008
- [7] Frederick Jelinek, Statistical Methods for Speech Recognition (Language, Speech, and Communication), 1998
- [8] Zheng-Hua Tan and Børge Lindberg, Automatic Speech Recognition on Mobile Devices and over Communication Networks (Advances in Pattern Recognition), 2008