대화식 휴대용 영어학습기 개발

Development of Portable Conversation-Type English Leaner

유 재 택*, 윤 태 섭**
(Jae-Tack Yoo, Tae-Seob Yoon)

Abstract - Although most of the people have studied English for a long time, their English conversation capability is low. When we provide them portable conversational-type English learners by the application of computer and information process technology, such portable learners can be used to enhance their English conversation capability by their conventional conversation exercises. The core technology to develop such learner is the development of a voice recognition and synthesis module under an embedded environment. This paper deals with voice recognition and synthesis, prototype of the learner module using a DSP(Digital Signal Processing) chip for voice processing, voice playback function, flash memory file system, PC download function using USB ports, English conversation text function by the use of SMC(Smart Media Card) flash memory, LCD display function, MP3 music listening function, *etc.* Application areas of the prototype equipped with such various functions are vast, *i.e.* portable language learners, amusement devices, kids toy, control by voice, security by the use of voice, *etc.*

Key Words : voice recognition, embedded systems, English learner

1. 서 론

우리나라 사람들의 대부분이 정규 교육과정에서부터 시작하여 회사원이 된 후에도 영어 학원, 학습지, 해외연수 등으로 영어 학습을 계속하나 영어회화의 능숙도가 미미하고 지속적인 영어사용이 부족하다. 이의 해결을 위하여 컴퓨터 및 정보처리 기술을 활용하여 누구나 편리하고, 효율적으로 영어 학습을 할 수 있게 해주는 대화식 휴대용 영어 학습기의 개발이 요구된다. 컴퓨터 및 정보처리 기술을 이용한 영어학습기의 핵심 필수 기술은 임베디드 환경의 음성인식 및 합성모듈이다. 본 논문에서는 음성인식 DSP(Digital Signal Processor)를 사용하여 음성인식 및 합성기술을 활용할 수 있는 소형 음성인식 및 합성 모듈 prototype을 제작했다. 음성인식 외에 음성녹음 및 재생, FM 라디오, MP3음악 감상, 기능을 포함한다. 음성인식은 전용 DSP를 사용하고, 음성 녹음 및 재생은 MP3 칩을 사용하여 음질을 높였다. 플래시 메모리에 다양한 영어학습 교재를 저장할 수 있고, USB 포트를 통하여 PC에서 인터넷으로 교재를 다운로드 할 수도 있다. 화자 독립 및 화자종속 방법을 모두 이용할 수 있으므로, 발음교정은 화자독립 방법을 이용하고 기기조작 명령어에는 화자 종속 방법을 적용하여 사용자 편리성을 높였다.

* 유재택: 안양대학교 전기전자공학과

** 윤태섭: (주)하이피크 대표이사

2. 음성 인식 및 합성

2.1 음성 인식

음성 인식은 일반적으로 마이크를 통하여 얻어진 음향 신호를 단어나 단어 집합 또는 문장으로 변환하는 과정을 말하며, 인식된 결과는 데이터 입력, 문서 준비 등의 응용 분야에서 최종 결과로 사용될 수 있고, 음성 이해와 같은 분야에는 언어 처리과정의 입력으로 사용될 수 있다. 음성 인식은 응용 분야와 사용기술에 따라 여러 가지 종류가 있는데 특정 사용자의 음성만을 인식하는 화자종속 음성인식과 여러 사람의 음성을 인식하는 화자독립 음성인식으로 구분된다. 화자종속 음성인식은 화자독립 음성인식에 비해 인식률이 높아 실용화하기에 유리하다. 또한, 음성인식의 단위에 따라 고립단어 및 연속음성 인식으로 나눌 수 있다. 고립단어 음성인식은 짧은 음성명령이나 간단한 음성제어 등에 주로 사용된다. 고립단어 음성인식은 인식률이 높고 구현하기 간단해 널리 이용되고 있으나 사용자가 이용하기 불편하다는 것이 단점이다. 연속음성 인식은 문장을 인식하기 때문에 사용자가 단어 단위로 끊어 발음하지 않아도 된다. 음성 인식 시스템은 보통 마이크로 입력된 음성신호를 표본 검출 후, 신호 분석을 거친다. 이 신호 분석 시 인간의 청각필터 특성을 감안하기도 한다. 그 결과는 특징벡터라 불리는 숫자 열을 구성한다. 특징벡터는 음향분석(Acoustic Analysis)을 거치게 되는데 이 때 두 가지의 음향모델(Acoustic Model)을 이용할 수 있다. 하나는 인식대상 단어, 음소 등 어떤 특징벡터를 모델로 설정하고 이와 얼마나 유사한 가 비교(Match)하는 방법과 또 다른 한 가지는 상태열(State Sequence)을 비교대상으로 쓸 수가 있

다. 상태열은 복잡한 모델을 표시 할 수가 있어 간단한 단어 인식을 제외하고는 대부분 이를 이용한다. 음향모델로는 단어, 음절, 반음절, 음소 등 다양한 단위를 사용할 수 있다.

2.2 음성 합성

음성합성은 음성분석 과정을 통해 압축되어 저장되었던, 음성을 구성하는 요소 변수들을 다시 조합하여 원래의 음성과 똑같은 음색을 갖는 소리를 재생시키는 것이다. 기계적인 장치나 전자회로 또는 컴퓨터를 이용하여 자동으로 음성파형을 생성해내는 것으로 정의할 수 있다. 음성합성 기술은 실제 응용 방식에 따라 크게 두 가지로 구분되는데 제한된 어휘 개수와 구문구조의 문장만을 합성하는 제한 어휘 합성 또는 자동음성응답 시스템((ARS: Automatic Response System)과 임의의 문장을 입력받아 음성 합성하는 무제한 어휘 합성 또는 문서-음성 변환(TTS: Text to Speech) 시스템이 있다.

3. 시스템 설계

설계한 영어 학습기의 기능은 화자종속 및 화자독립 음성인식 기능, 대화·듣기·따라하기(발음교정), 동화·동요·팝송(캡션기능 포함), 교재 인터넷 다운로드 및 SMC를 사용하여 교재를 교체하면 다양한 언어학습이 가능하도록 설계하였다. MP3 음악감상, FM 라디오 기능, 음성녹음, 알람시계 기능, USB 컴퓨터 마이크 및 스피커, 절전모드 등의 부가기능도 구현을 하였다. 마이크를 통하여 입력된 음성 신호는 ADC (Analog-Digital Converter)를 이용하여 변환하고, 음성인식 DSP를 이용하여 단어별 음성인식을 하고, 화자독립은 단어를 matching 시켜 인식을 하며, 화자 종속은 사용자의 음성을 녹음하여 음성명령으로 기기를 조작할 수 있다. 아래의 그림 1은 전체시스템구의 블록 다이어그램을 보여준다. 이 그림에서는 음성인식 DSP 와 MP3 encoder/decoder chip 응 사용하였음을 보여주며 이외에 flash memory, SRAM, ADC, DAC, LCD등의 사용을 보여준다.

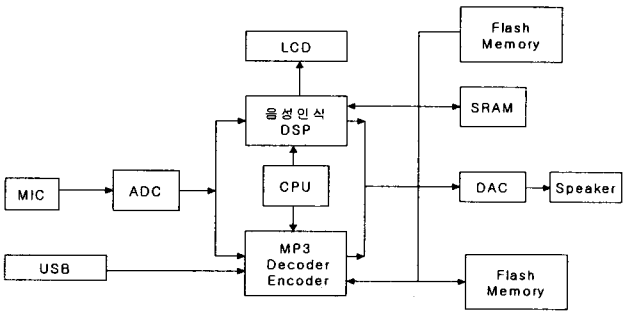

그림 1. 전체 시스템 구성도

3.1 음성 인식 시스템

음성인식 시스템은 음성인식 DSP를 이용하였다. 음성 최적화를 위한 음성인식 DSP로는 Sensory RSC-4x Series의 RSC-4000을 사용하였다. RSC-4x는 97%이상의 화자독립 음성인식과 99%이상의 화자종속 음성인식의 정확성을 가진다. 그림 2는 음성인식 시스템을 보여주며, 음성인식의 동작수행을 위해 내부에 8-bit micro controller, pre-amplifier, twin-DMA(Direct Memory Access) units, vector accelerators, hardware multiplier, 16-bit ADC, 10-bit DAC, 3 timers, 4.8 Kbytes RAM으로 구성되어 있어 사용이 편리하다.

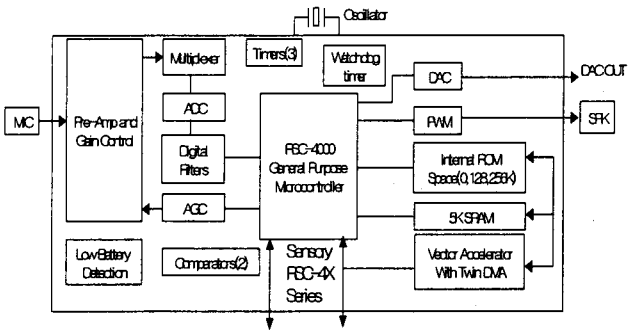

그림 2. 음성인식 시스템

3.2 MP3 Encoder/decoder 시스템

아래 그림3은 MP3 Decoder/Encoder를 위한 Skylark CODEC을 이용한 시스템을 보여준다. Skylark은 MP3 Encoder, Decoder, DMA Controller, USB/UART Controller, 80C51, Media Interface(MMC/SMC/Flash)등을 하나의 chip 상에 구현한 Audio/Voice CODEC이다. Hardwired Logic으로 구현된 저전력 MP3 CODEC Engine에 의해서 Audio/Voicedml 녹음 및 재생이 가능하며, 고속의 DMA를 통해 Skylark에 내장된 USB/UART, Media Interface등의 기능을 수행할 수 있게 해준다. USB/UART는 PC와의 Interface를 제공하여 MP3 Audio/Voice File을 Media Memory로 upload 또는 download할 수 있게 해준다. 내장된 8051 MCU에 의해서 Data File 관리 및 Interface를 수행할 수 있다.

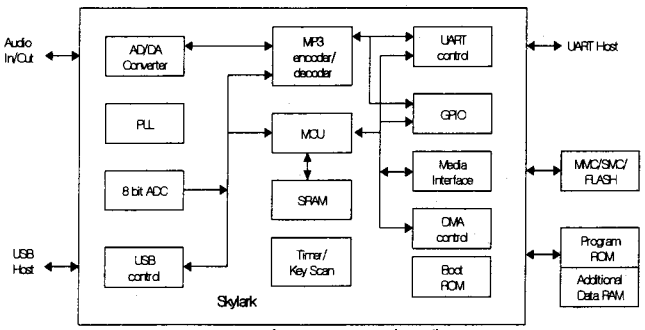

그림 3. MP3 시스템

4. Hardware Prototype

아래 그림 4는 설계된 하드웨어의 전체 구성도를 보여준

다. 영어노래, 동요, 동화 듣기, 음성녹음, FM Tuner 제어, FM 녹음 등은 Skylark을 이용하여 구현하였으며, 음성인식 및 음성 합성 기능은 Sensory RSC-4000을 사용하여 구현하였다. NAND Flash 및 SMC(Smart Media Card) access는 Skylark을 활용한다. 음성 입력 및 출력은 높은 입질을 위하여 MP3 칩을 사용하고, 음성인식 칩은 음성인식만을 수행한다. 음성인식 칩과 MP3 칩은 플래시 메모리 데이터 및 입출력 장치를 공유하기 위해 서로 통신을 한다.

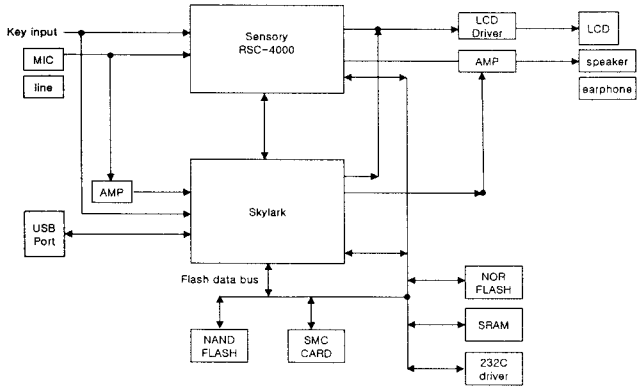

그림 4. 하드웨어 구성도

아래 그림 5는 PCB(Printed Circuit Board)에 개발하여 완성한 음성인식 module을 보여준다. 하단의 LCD에는 text를 출력하여 사용자의 발음 교정 및 대화 관련한 교정시에 도움이 되도록 하였으며, 그 우측에 SMC를 삽입하여 교육용 영어교재 text file을 사용할 수 있도록 하였다.

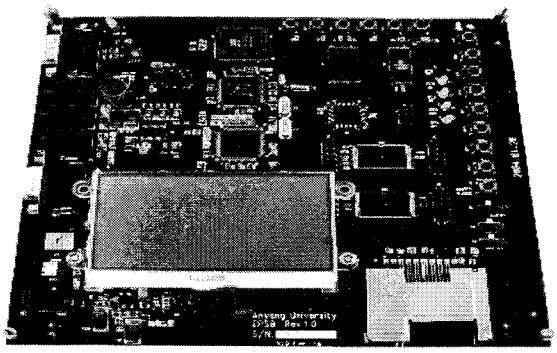

그림 5. 완성된 음성인식 모듈

4. 결 론

본 논문에서는 음성인식 및 합성기술을 활용하여 소형 음성인식 및 합성 모듈 prototype을 설계한 결과를 보여주었으며, 플래시 메모리에 음성합성음을 저장하여 재생하는 기능, 음성인식 DSP와 MP3 chip간의 통신기능, USB를 이용한 PC와의 통신, 플래시 메모리 파일 시스템, 영어회화 교재사용을 위한 SMC(Smart Media Card)플래시 메모리, LCD 디스플레이, 부가기능으로 MP3 encoder/decoder chip을 사용하여 MP3 음악 감상 기능 등을 구현한 결과를 보여 준다. PCB module 구현 및 시험결과 기본적인 system 작동이 원활하게 되고 있음을 확인하였다. 본 시스템의 대화형 학습기능을 충분히 구현하기 위하여서는 PC software를 사용하여 본 hardware에 download되어 사용될 실행 file을 확장 개발하여야 하며, 본 시스템을 이용한 휴대용 기기의 완성을 위하여 본 prototype의 hardware 중 핵심부분만 취하여 다시 제작하여야 한다.

참 고 문 헌

[1] 오영환, "음성언어정보처리", 홍릉과학출판사
[2] 김영택 외, "자연언어처리", 생능출판사
[3] 박경범, "음성의 분석 및 합성과 그 응용", 그린출판사
[4] 박섭형, "음성 웹 애플리케이션 구축을 위한 VoiceXM", 한빛미디어
[5] Yoshua Bengio, "Networks for Speech Sequence Recognition", Tomson
[6] A.Syrdal, R.Bennett, S.Greenspan, "Applied Speech Technology", CRC
[7] RSC-4X Data sheet, SENSORY
[8] Skylark Specification, Enhanced Chip Technology Inc.