

음성인식 기술을 이용한 대화식 언어 학습기 개발

Development of Language Study Machine Using Voice Recognition Technology

유재택*, 윤태섭**

(Jae-Tack Yoo, Tae-Seob Yoon)

Abstract - The best method to study language is to talking with a native speaker. A voice recognition technology can be used to develop a language study machine. SD(Speaker dependant) and SI(speaker independant) voice recognition method is used for the language study machine. MP3 Player, FM Radio, Alarm clock functions are added to enhance the value of the product. The machine is designed with a DSP(Digital Signal Processing) chip for voice recognition, MP3 encoder/decoder chip, FM tuner and SD flash memory card. This paper deals with the application of SD and SI voice recognition, flash memory file system, PC download function using USB ports, English conversation text function by the use of SD flash memory, LCD display control, MP3 encoding and decoding, etc. The study contents are saved in SD flash memory. This machine can be helpful from child to adult by changing the SD flash memory.

Key Words : voice recognition, language study, MP3

1. 서 론

언어학습의 가장 좋은 방법은 사람들과 대화를 하는 것이다. 그래서 많은 사람들이 언어학습을 위해 어학연수, 유학 등으로 외국을 방문하거나, 원어민이 강의하는 학원을 다닌다. TV 강의나 카세트테이프나 MP3 player 등으로 강사의 발음을 듣고 따라하는 방법은 흥미가 적고, 자신을 발음 정확도를 모른다. 최근 급속히 발전하는 음성인식 기술을 이용하여 대화식 언어학습기를 개발하면 적은 비용으로 시간과 장소에 관계없이 효과적으로 언어를 학습할 수 있다.

언어학습기 위주의 단일기능 제품은 학습하는 동안만 사용하므로 활용도가 낮으므로 MP3 player, FM 라디오, 알람시계 등의 기능을 추가하여 활용도를 높인다. 음성인식 전용 DSP 칩은 음성인식 기능만을 수행하고, 음성의 출력, 음성녹음 등의 기능은 MP3 칩에서 수행하여 깨끗한 음성을 구현하였다. 음성인식 칩과 MP3 칩은 별도의 CPU를 내장하고 있으므로 칩 사이의 통신은 RS232 방식을 사용하였다.

언어학습 교재는 SD 플래시메모리 카드에 저장하여 메모리를 교체하면 다양한 학습교재를 사용하여 한 개의 학습기로 어린이부터 성인까지 사용할 수 있도록 하였다.

* 유재택: 안양대학교 전기전자공학과 교수

** 윤태섭: (주)하이피코 대표이사

2. 음성 인식

일반적으로 음성 인식은 마이크를 통하여 입력된 음성신호를 단어나 문장으로 변환하는 과정을 말하며 보통 PC에서 Dictation 기능이라 한다. Dictation 기능을 효과적으로 수행하기 위해서는 음성인식엔진에서 제공하는 표본문장을 읽어 사용자의 음성을 컴퓨터가 분석하여야 인식율이 높아진다.

임베디드 환경에서의 음성인식은 프로세서의 처리능력, 메모리, OS 등의 한계로 PC와 같은 Dictation 기능을 수행하기는 어렵고, 특정한 목적에 사용되는 단어 또는 문장을 DB에 입력해 놓고, DB의 저장된 단어나 문장에서 마이크로 입력된 음성과 일치하는 것을 찾는다. 화자종속 방법에서는 사용자의 음성을 녹음하여 DB를 만들지만, 화자독립 방법에서는 아이, 어른, 남자, 여자 등 200명 이상의 음성을 녹음하여 분석하고 모델링하여 DB를 만든다. 화자독립 DB를 작성하는 작업은 많은 시간과 비용이 소모된다. 최근에는 텍스트를 입력하면 DB를 생성하는 소프트웨어를 사용하면 쉽게 화자독립 DB를 쉽게 생성할 수 있다. 본 연구에서는 미국의 Sonymory사에서 개발한 T2S이라는 우수한 DB 작성 tool을 사용하였다.

음성인식에서는 인식에 사용되는 DB의 크기가 작을수록 인식율이 높아진다. 언어학습 교재는 수백개의 문장을 포함하므로 수백개의 문장에서 한문장을 찾으려면 인식율이 낮아진다. 교재를 20개 정도의 문장을 한 단원으로 분류하고 20개의 문장중에서 사용자의 음성과 일치하는 문장을 찾으면 인식율을 높일 수 있다. 한 단원에서는 발음이 비슷한 문장

을 포함하면 서로 혼동하여 인식이 낮아지므로 발음이 비슷한 문장을 같은 단원에 포함하지 않도록 하는 것이 좋다.

음성합성은 컴퓨터가 말을 하는 것과 같은 것으로 문장을 입력하면 음성파형을 생성한다. 음성합성은 임베디드 환경에 CPU의 처리능력, 메모리 용량, OS 등의 제한 때문에 자연스런 음성파형을 생성하기는 어렵다. 어학학습기에서는 제조사가 공급한 교재만 학습할 수 있지만, 음성합성시스템은 사용자가 필요한 교재를 입력하면 원어문과 비슷한 음성파형을 생성한다. 국내에서 개발한 음성합성 솔루션인 Natural English는 PC에서 음성파형을 .wav 또는 MP3 포맷으로 생성한다. PC에서 들을 수 있고, MP3 포맷 음성파일은 학습기로 다운로드 받아 동기학습에 사용할 수 있다.

3. 시스템 설계

시스템의 기능 구현을 위하여 Sensory 사의 음성인식 전용 DSP 칩 RSC-4128, ECT 사의 MP3 encoder/decoder 칩 Skylark-A, FM Tuner 모듈 등을 사용하였고, 플래시 메모리 카드, USB 통신, LCD, 데이터 및 프로그램 저장용 플래시 메모리, 입력 키, 전원공급 부분 등으로 구성하였다. 그림1은 전체시스템의 블록 다이어그램을 보여준다. 마이크로 입력된 음성신호는 음성 코덱을 거쳐 음성인식 칩과 MP3 칩으로 입력된다. 음성인식 칩에는 코덱이 내장되어있다. MP3 칩은 음성을 녹음할 때 사용되며 MP3 포맷으로 저장하므로 메모리를 절약하고 높은 음질을 유지할 수 있다. FM 라디오 부분은 전원불안정으로 인한 잡음영향을 적게 받도록 다른 부분과 분리된 전원을 공급하도록 하였다.

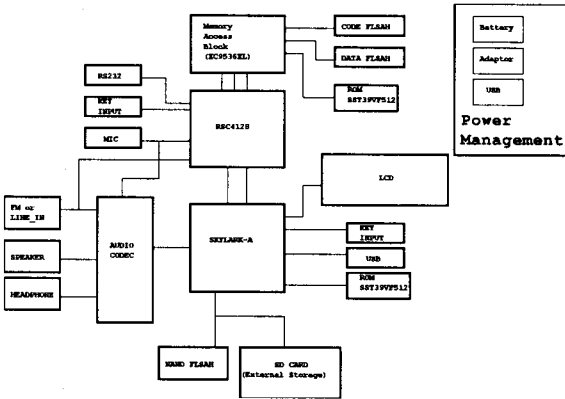


그림 1. 전체 시스템 구성도

3.1 음성 인식 모듈

음성인식 모듈은 미국 Sensory 사의 RSC-4x Series의 RSC-4128을 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. Sensory사의 음성인식 DSP RSC-4128은 언어학습 뿐만 아니라 완구, 음성제어 시스템 등에 사용할 수 있다. 음성인식 소프트웨어는 미국 Phytion 사의 C 컴파일러를 사용하여 개발하였다.

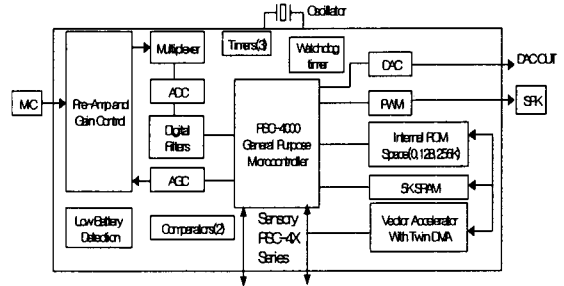


그림 2. 음성인식 모듈

3.2 MP3 Encoder/decoder 모듈

아래 그림3은 MP3 Decoder/Encoder를 위한 국내의 ECT 사의 Skylark-A CODEC을 이용한 시스템을 보여준다. Skylark-A는 MP3 Encoder, Decoder, DMA Controller, USB/UART Controller, 80C51, Media Interface (MMC/SMC/Flash)등을 하나의 chip상에 구현한 Audio CODEC이다. Hardwired Logic으로 구현된 저전력 MP3 CODEC Engine에 의해서 Audio/Voice을 녹음 및 재생이 가능하며, 고속의 DMA를 통해 Skylark-A에 내장된 USB/UART, Media Interface등의 기능을 수행할 수 있게 해준다. USB/UART는 PC와의 Interface를 제공하여 MP3 Audio/Voice File을 Media Memory로 upload 또는 download 할 수 있게 해준다. 내장된 8051 MCU에 의해서 Data File 관리 및 Interface를 수행할 수 있다. MP3 펌웨어는 Keil C를 사용하여 개발하였다.

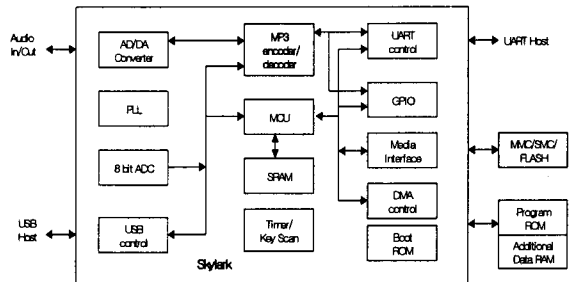


그림 3. MP3 모듈

4. 시제품 개발

아래 그림 4는 설계된 하드웨어의 전체 구성도를 보여준다. 영어노래, 동요, 동화 듣기, 음성녹음, FM Tuner 제어, FM 녹음 등은 Skylark-A를 이용하여 구현하였으며, 음성인식 및 음성 합성 기능은 Sensory RSC-4128을 사용하여 구

현하였다. NAND Flash 및 SD(speaker dependent)는 Skylark-A를 활용하였다. 음성 입력 및 출력은 높은 음질을 위하여 MP3 칩을 사용하고, 음성인식 칩은 음성인식만을 수행한다. 음성인식 칩과 MP3 칩은 플래시 메모리 데이터 및 입출력 장치를 공유하기 위해 서로 RS232 방식 통신을 한다.

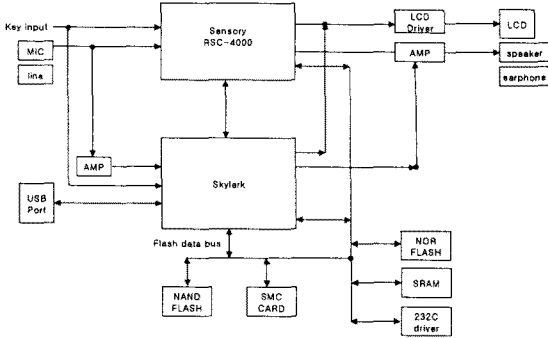


그림 4. 하드웨어 구성도

아래 그림 5는 PCB에 부품을 조립하여 제작한 시제품의 음성인식 학습기 타겟보드를 보여준다. 보드 크기는 12cm×13cm 이다. 상단의 LCD에는 text를 출력하여 사용자의 발음 교정 및 대화 관련한 교정시에 도움이 되도록 하였으며, 그 우측에 SD 카드를 삽입하여 교육용 영어교재 text file을 사용할 수 있도록 하였다.

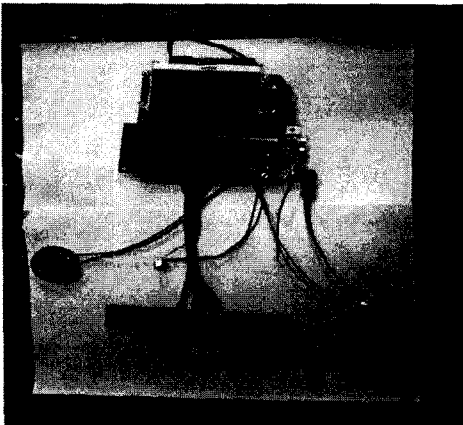


그림 5. 완성된 음성인식 학습기 타겟보드

4. 결 론

본 논문에서는 음성인식 기술을 이용한 대화식 언어학습기 개발 및 타겟보드를 보여주었으며, 플래시 메모리에 음성합성을 저장하여 재생하는 기능, 음성인식 DSP와 MP3 chip간의 통신기능, USB를 이용한 PC와의 통신, 플래시 메모리 파일 시스템, 영어회화 교재사용을 위한 SD 카드 플래시 메모리, LCD 디스플레이, 부가기능으로 MP3 encoder/decoder

chip을 사용하여 MP3 음악 감상 기능 등을 구현한 결과를 보여 준다. PCB module 구현 및 시험결과 음성인식, MP3 Player, FM 라디오 등의 기능이 원활하게 작동되고 있음을 확인하였다.

본 시스템의 대화식 학습기능을 충분히 활용하기 위해서는 다양한 학습 콘텐츠를 별도로 제공하여야 하며 또한 활용성을 높이기 위해서는 예쁘고 실용적인 케이스를 디자인하여 친숙도를 높여 주어야 한다. 한국, 중국, 일본 등에서는 영어 학습을 많이 시키고 있으며, 미국, 영국, 독일 등 전세계 모든 국가에서 외국어 학습을 하므로 향후 이러한 기법의 적용을 확장하면 영어, 중국어, 일본어, 스페인어 등 많은 언어학습에 적용할 수 있다.

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, "Neural 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.
- [9] 차영배 "One-Chip Micro Computer 8051"
- [10] M. Morris Mano " Dital Design"