

# 음성인식을이용한영어학습기구현 Realization of the Language Instructor Using Speech Recognition

신승식(Seung-sik shin)<sup>1)</sup> 전형준(Hyung-Joon Jeon)<sup>2)</sup> 정찬수(Chan-Soo Chung)<sup>3)</sup>  
유봉선(Bong-Sun Yu)<sup>4)</sup> 조정현(Kyung-Hyun Cho)<sup>5)</sup> 강창수(Chang-Soo Kang)<sup>6)</sup>

韓國컴퓨터産業敎育學會 論文誌 '2004. 9. Vol 5., No 9.요 약

본 논문에서는 Conversay 음성인식엔진 사용법을 근간으로 한 Conversay SDK로 3개의 시나리오를 바탕으로 프로그램화하여 음성인식을 이용한 어학 학습기를 구현하였다. 이를 위하여 리눅스를 탑재하였고, 마이크로 프로세서로 Intel의 StrongARM SA1110을 사용하여 구현하였으며, 기능으로는 영어 발음/ 문장 인식 기능, 영어 대화 기능과 같은 기본기능 외에 음성을 이용한 시간 조정, 알람 기능, Test 기능, 학습 Check 기능 등을 추가하였다.

## ABSTRACT

This study will realize the learning device of foreign languages using voice recognition that is programmed based on three scenarios by the Conversay SDK. This device is embedded with Linux and Intel's Strong ARM SA1110 as a microprocessor. Also it is equipped not with basic functions such as pronunciation/phrase recognition function, conversation function etc.,but also with additional functions using user's voices such as timer function, alarming function, test function, learning check function etc.

논문접수 : 2004. 11. 15.

심사완료 : 2004. 12. 15.

- 1) 정회원 : 인덕대학 인터넷 정보통신계열
- 2) 정회원 : LG전선 광통신연구소 책임연구원
- 3) 정회원 : 숭실대학교 전기제어학부
- 4) 정회원 : 인덕대학 인터넷 정보통신계열
- 5) 정회원 : 광운대학교 대학원 전자공학과 박사과정수료
- 6) 정회원 : 유한대학 전자정보과

### 1. 서론

오늘날 전기, 전자, 컴퓨터 기술들이 발달하면서 전자기기들은 미리 정해진 특정 기능을 수행하기 위해 컴퓨터 하드웨어와 소프트웨어가 조합된 임베디드 시스템을 이용한 것이다. 이에 본 연구에서는 오늘날의 국제화, 세계화 추세에 발맞추어, 기존 영어 학습 시스템의 문제점을 보완하기 위하여, 임베디드 시스템 기반으로 언제, 어디서나, 누구나, 원할 때마다 외국인을 직접 만난 것같이 영어회화를 자유롭게 공부할 수 있는 음성인식을 이용한 어학 학습기를 구현하고자 한다. 임베디드 시스템의 OS는 시스템의 규모가 커짐에 따라 Multi Tasking 등과 같은 복잡한 기능을 요구하며, TCP/IP, GUI, Audio, Video 등 Network이나 multimedia가 시스템의 기본으로 자리잡아가고 있다. 또한 임베디드 시스템의 특성상 가능한 실시간 OS이어야 하며, 이를 만족시켜야만 하는 시스템도 늘어나고 있는 추세이다. 현재 임베디드 시스템을 운용하기 위해 WinCE, pSOS, VxWorks, VRTX, Palm OS 등 다양한 OS가 사용되고 있으며, 이러한 OS 중 본 연구에서는 Linux를 사용하였다.

### 2. 시스템 구성 및 Target Board

구현된 "음성인식을 이용한 어학 학습기"의 시스템 블록은 그림 2.1과 같다.

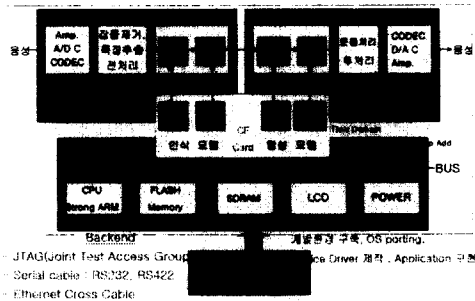


그림 2.1. 구현한 어학학습기의 System Block Diagram

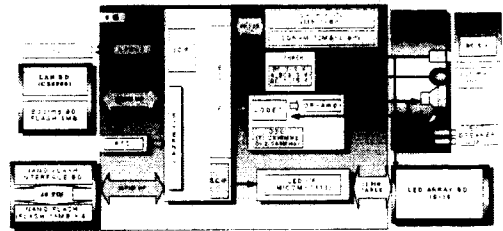


그림 2.2 어학학습기의 Target Board 구성도

어학 학습기의 Target Board 사양 Table 2.1 과 Target Board 구성도 그림. 2.2를 보면 알 수 있듯이, CPU는 저전력 고성능의 Intel StrongARM SA1110을 사용하였고, Flash 메모리는 Intel Strata Flash Memory 4M, 64M (또는 32M)를, 그리고 SRAM은 32M를 사용하였다. 보드의 구성은 기본적인 사양에 충실하게 디자인하였다. 먼저 개발환경으로 모니터링과 Serial downloading이 가능하도록 Serial Port 3개로 구성하였으며, network과 연동할 수 있도록 CS8900 Ethernet Controller를 사용하여 network 및 tftp 기능을 사용할 수 있도록 또한 구성하였다. 또 JTAG Port를 통하여 H/W Test 및 Flash Memory에 데이터를 fusing할 수 있도록 구성하였을 뿐만 아니라, SA1110의 모든 기능을 사용 가능하도록 하기 위해 확장용 connector를 사용하여 Board를 구성하였다. 그리고 LDM과 GPIO S/W를 이용하여 디바이스 드라이버를 제작하여 테스트가 용이하도록 Board를 설계하였다.

표 2.1 어학학습기 Target Board의 사양

CPU	Intel StrongARM SA-1110 206MHz
Memory	Flash Memory 4Mbyte (2M x 16bit)
	SRAM 32Mbyte (8M x 16bit x 22개)
Display	LDM Module 256(16x16), RED, GREEN, GRAMME
	Size 64mm x 64mm
Speaker	대형 스피커 (8ohm, 0.5W), Speaker Jack
NIC	대형용 마이크로, NIC Jack
Real Time Clock	RTC 전용 IC 사용 Li-Ion (RCS, 7V x 90mAh) RTC Back-up 용
외부메모리 (CF Card)	Compact Flash Card Type I (32MB/64MB)
전원	AC/NIC ADAPTER (5V-5A, 5V-1A Dual 출력)
Board	Ethernet CS8900
	Serial 3 Port
	USB Type B Port
	JTAG
	GPIO
Connector	SA1110 확장용

### 3. 어학 학습기의 S/W 구성

#### 3.1. Cross 개발환경 구축 및 BLOB( Boot Loader Object)

H/W 제작이 끝나면 Cross 개발환경을 구축 하여야 하는데, 본 연구에서는 StrongARM SA1110에 사용한 ARM tool chain을 구성하였다. 일반적으로 부트로더라하면 일반 x86리눅스에선 LILO를 많이 사용한다. LILO란 Linux Loader로써 MS-Windows나 NT, 리눅스 등 다른 OS를 선택적으로 부팅할 수 있도록 하는 기능을 제공한다. LILO는 하드디스크의 MBR에서 동작이 되는 프로그램으로 OS가 실행할 수 있도록 점프하는 기능을 수행한다. 본 연구에서 사용된 BLOB 부트 로더는 flash 0 블록에서 실행되고, 다음과 같은 여러 가지 다양한 기능들을 수행한다.

- ① 하드웨어의 초기화
- ② Linux Booting
- ③ Kernel or ramdisk Download
- ④ Downloading한 kernel 및 Image를 flash에 Writing

#### 3.2. JTAG을 이용한 Flash Memory Fusing

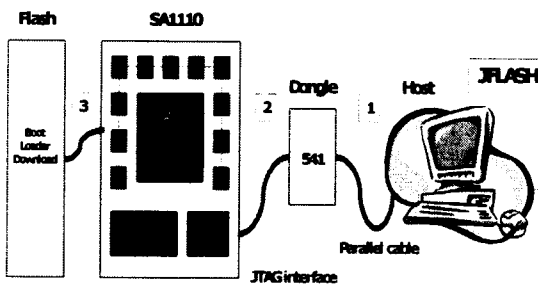


그림 3.1 JTAG을 이용한 flash fusing 방식

그림 3.1과 같이 host system으로부터 Target board에 있는 strata flash에 boot loader를 쓰기 위해서는 위의 그림과 같은 구성이 필요하다. 먼저 host system의 구성부터 살펴보면 host system에서 동작하는 jflash라

는 프로그램이 필요하다. 이 jflash는 parallel port를 이용하여 target board에서 필요로 하는 JTAG 신호를 생성한다. Host system에서 생성된 JTAG 신호는 parallel cable을 통해 dongle에 전달된다. Dongle은 TTL 74HCT541을 사용하여 구현하였으며, 이 기능은 단지 host system의 parallel port에서 발생한 5V 전압을 SA1110에 적합한 3.3V 전압 레벨로 변환하는 것이다. Dongle을 통해서 전달된 JTAG 신호 중 TMS와 TCLK는 TAP에 전달되어 JTAG의 state machine을 결정하고, TDO와 TDI는 JTAG의 명령의 상태에 따라 bypass register, boundary scan cell, ID register 등의 입력과 출력 부분에 연결된다. SA1110의 JTAG을 통해서 flash 메모리가 필요로 하는 bus timing을 발생하여 flash에 전달한다. host system의 shell에서 jflash blob을 입력하면 blob의 binary code가 flash 메모리의 0번지부터 fusing을 한다. 메모리에 쓰기 전에 기본적으로 쓰고자 하는 0번 block을 지우고, blob을 fusing한 후, 에러 없이 fusing 되었는지를 검사하기 위해 검증을 한 후에 이상이 없으면 blob이 정상적으로 loading을 완료하게 된다. 이 상태에서 serial 통신 환경의 설정에 이상이 없다면, 정상적으로 blob이 동작하는 것을 볼 수 있다.

#### 3.3. Application 프로그램

본 연구에서는 본 개발 목적에 맞는 음성인식엔진을 선택하기 위해 지금까지 개발되어 사용되고 있는 음성인식엔진들을 상호 비교 검토한 결과, Conversay의 영어 음성인식엔진을 선택하였다. 국내 음성인식엔진에는 VoiceWare, VoiceTech, D&M 엔진 등이 있으나, 본 연구에서는 음성인식 엔진 선택시 고려 대상에서 제외시켰다. 그 이유는 본 연구에서는 음성인식을 하는 어학학습기, 특히 영어 어학학습기를 개발하는 것이 목적이므로, 영어 문장들 중 연음 처리를 해야 하는 문장들을 영어 발음 그대로 한국말로 처리하기가 어려운 경우가 많이

발생하기 때문이다. 영어 중심의 음성인식엔진은 Nuance, SpeechWorks, L&H, Conversay 등등이 있으나, 이 중 Embedded System 기반으로 음성인식을 이용하여 영어 어학학습기 구성시, 엔진의 용량과 성능 및 엔진을 응용하고자 할 때 support해줄 수 있는 Engineer의 한국 존재 여부 등을 고려하여 본 연구에서는 Conversay 음성인식을 선택하였으며, 그 결과 Conversay SDK(Software Development Kit)로 응용 프로그램을 3.2.4-2)절과 같은 시나리오로 짜서 사용자의 음성을 음성인식을 할 수 있는 영어 어학학습기를 개발하였다.

### 3.4.1. 음성인식엔진

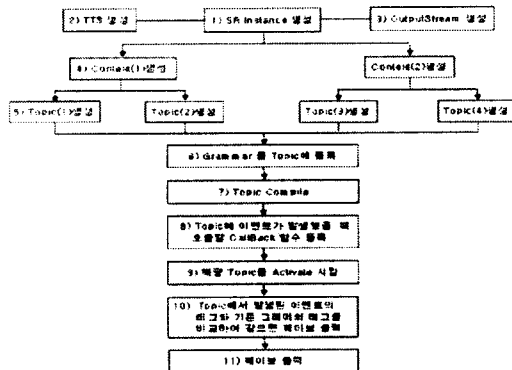


그림 3.2 Conversay 음성인식엔진 활용 방법

Conversay 음성인식엔진을 사용하는 방법을 간단히 그림. 3.2에 도시하였고, 그 과정을 다음에 나타내었다.

\* Grammar 등록 : Grammar 등록은 인식할 음성을 텍스트로 미리 등록하는 부분으로, Topic을 Activate하기 이전에만 생성해주면 된다.

LPTSTR

study\_grammar=TEXT("<study>::~=PREVIOUS\_MODE:PM)

1) SR (Speech Recognition) : 음성 인식을 시

작하기 위해 음성인식 객체를 Open해주어야 한다.

CVACreateInstance(CLSID\_CVASRInstance, NULL, 0,

IID\_ICVASRInstance, (void \*\*)&m\_pSR);

2) TTS(Text-To-Speech) : 문자를 음성으로 변환해주는 객체이다. 문자를 입력하고 OutputStream으로 출력을 보내면, 컴퓨터 합성음이 출력된다.

CVACreateInstance(CLSID\_CVATTSTInstance, NULL, 0,

IID\_ICVATTSTInstance,

(void \*\*)&m\_pTTS);

3) OutputStream : TTS나 웨이브를 출력해주는 객체이다.

CVACreateInstance(CLSID\_CVAOutputStream, NULL, 0,

IID\_ICVAOutputStream,

(void \*\*)&m\_pOstream)

4) Context : SR로 음성인식을 위한 객체를 생성하고 난 후, 여러 개의 Topic을 묶을 수 있는 Context를 생성한다. Context를 생성하고 난 후에는 SR객체에 그 Context를 연결해야 한다. Context는 필요에 따라 하나 이상의 객체를 생성할 수 있다.

CVACreateInstance(CLSID\_CVAContext, NULL, 0, IID\_ICVAContext,

(void \*\*)&m\_pMainContext);

5) Topic : 실제적인 문장을 받아들이는 객체이다. Topic도 여러 개의 토픽을 생성할 수 있다.

CVACreateInstance(CLSID\_CVATopic, NULL, 0, IID\_ICVATopic,

(void \*\*)&m\_pStartTopic);

6) Grammar를 Topic에 등록 : 설정된 grammar를 적당한 Topic에 등록한다.

m\_pStartTopic->SetGrammar(start\_grammar, GRAM\_FMT\_BNF);

7) Compile : Topic에 등록된 grammar가 제대로 등록이 되었는지 확인하기 위해 컴파일한다. m\_pStartTopic->Compile();

8) SetEventCallBack : Topic에 발생한 이벤트를 처리해 주는 함수를 선언하는 부분이다.

m\_pStartTopic->SetEventCallback(WORD\_EVENT, this, StartCall);

9) Activate/Deactivate : 등록된 Topic을 활성화할 것인지, 비활성화할 것인지를 결정한다.

m\_pStudyTopic->Activate();  
m\_pDictationTopic->Deactivate();

10) Event비교 : Topic에 등록된 grammar와 인식된 음성을 비교해서 처리하는 함수이다. 실제적인 처리부분으로, Tag(문장처리)처리와 Word(단어처리)처리 부분으로 나누어져 있으며, 사용 목적에 따라 적합한 방식을 선택 사용할 수 있다.

!\_tscmp(Msg->eventData.RecoWord.pszTag, TEXT ("H"))

11) SubmitFile : TTS나 웨이브 파일을 출력해 주는 함수이다. 보통 이벤트를 비교하여 인식한 음성과 Topic에 등록된 grammar가 일치하면, 웨이브를 출력시키는 함수이다.

m\_pOStream->SubmitFile(file, false);

3.4.2. Application 프로그램

앞절에서 설명한 Conversay 음성인식엔진 사용법을 근간으로 하여 Conversay SDK로 다음과 같은 시나리오 I, II, III를 각각 프로그램화하여 음성인식을 이용한 어학학습기를 구현하였다.

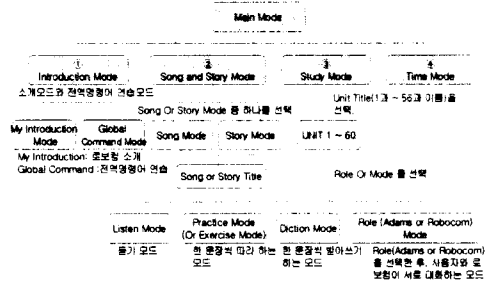


그림 3.3. 어학학습기 Scenario I

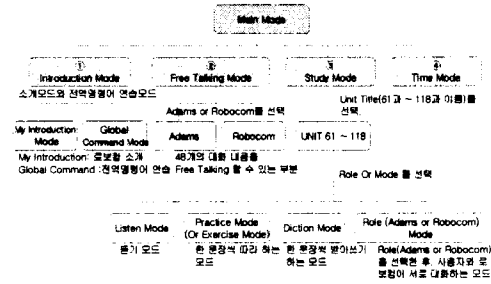


그림 3.4 어학학습기 Scenario II

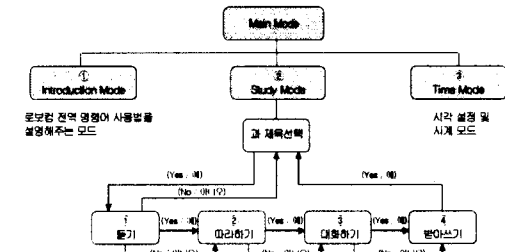


그림 3.5 어학학습기 Scenario III

4. 결론

개발한 Embedded system 기반의 음성인식을 이용한 어학학습기는 다음과 같은 기능을 갖고 있다.

- 1) 영어 발음, 문장 인식 기능
- 2) 영어 대화 기능(로봇과의 대화)
- 3) 문장 재생 및 반복기능

- 4) LDM 화면 Display 기능
- 5) 음성을 이용한 시간조정. 및 알람 기능
- 6) 구간 반복 기능
- 7) Track 찾기 기능
- 8) Test 기능 (최종적인 평가)
- 9) 학습 Check 기능

### 참고문헌

1. 박영환, "임베디드 시스템 & 임베디드 리눅스", (주)사이텍미디어, 2002. 6.
2. Jhon Lombardo, "임베디드 리눅스", 인터북, 2002.2.
3. 박철, "StrongARM(SA1110)을 이용한 board level porting 작업", www.kelp.or.kr, 2001.11.
4. 박철, "Howto 문서(4) - 커널 패치, 램디스크 만들기", www.kelp.or.kr, 2001.11.
5. 박철, "Howto 문서(2) - 디바이스드라이버", www.kelp.or.kr, 2001.11.
6. 박철, "CS8900 요약", www.kelp.or.kr, 2001.11.

신 승 식(Seung-Sik Shin)



1993년 2월 호서대학교 제어계측공학과  
1997년 2월 숭실대학교 대학원 공학석사  
2004년 8월 숭실대학교 대학원 공학박사

현재 인덕대학 인터넷.정보통신계열 정보통신전공 겸임교수

<주관심분야>

데이터베이스와 네트워크 통신, 임베디드 시스템, 백신 시스템

전 형 준



1993년 2월 광운대학교 전자공학과  
1995년 2월 광운대학교 전자공학과 대학원 공학석사  
2003년 2월 광운대학교 전자공학과 대학원 공학박사

2001년 2월~2004년 2월 삼성에스원 기술연구소

2004년 10월~ 현재 LG전선 광통신연구소 책임연구원  
<주관심분야>

M/W 및 RF 회로 설계, 근거리 무선통신 시스템

정 찬 수



1972년 2월 서울대학교 전기공학과  
1980년 2월 동대학원 공학석사  
1987년 2월 동대학원 공학박사  
현재 숭실대학교 전기제어학부 정교수  
<주관심분야>

적응 제어 시스템, 인공지능 시스템

유 봉 선



1984년 2월 성균관대학교 전자공학과  
1986년 2월 동대학원 공학석사  
1997년 2월 동대학원 공학박사  
현재 인덕대학 인터넷.정보통신계열 정보통신전공 정교수  
<주관심분야>

나노시스템, 광통신 시스템, 아날로그와 디지털 통신 시스템

조 경 현(Kyung-Hyun Cho)

1993년 2월 호서대학교 제어계측공학과 공학 학사  
1999년 8월 광운대학교 대학원 전자공학과 공학 석사

2003년 2월 광운대학교 대학원 전자공학과 박사과정 수료

<주관심분야>

변 복조 시스템, 광통신 시스템, 디지털 제어 시스템

강 창 수

1982년 2월 광운대학교 전자공학과  
1986년 2월 한양대학교 대학원 공학석사  
1992년 2월 광운대학교 대학원 공학박사  
1995년12월~1996년12월 Clemson University Post Doctor

1991년 3월 ~ 현재 유한대학 전자정보과 교수  
<주관심분야>

신경회로망, 반도체 회로설계, Semiconductor Reliability