음성인식 등의 복합기능을 가진 지능형 장난감의 소프트웨어 개발

박 상 훈° 한 상 훈 조 형 제
동 국 대 학 교 컴 퓨 터 공 학 과
{park77, hansh, chohj}@dongguk.edu

Software Development of an Intelligent Toy with Various Functions Including Speech Recognition

Sang-Hun Park Sang-Hoon Han Hyung-Je Cho
Dept. of Computer Science, Dongguk University

요 약

음성인식은 여러 분야에 적용될 수 있지만 지능형 장난감에 적용된 사례를 보면 다른 시스템에서 적용된 경우와 같이 높은 인식률이 요구된다. 하지만 음성인식의 기능만으로는 지능형 장난감의 기능이 다양성을 가지지 못한다. 음성인식기능 뿐만 아니라 다른 여러 가지의 기능을 가진 지능형 장난감의 소프트웨어를 개발하는 것이 다른 시스템과의 차별성을 두는 것이 된다.
본 논문에서는 이 Intelligent Toy에 내장될 음성인식 등의 여러 가지의 기능을 가진 Software를 구현하는 방법 및 그 결과를 제시한다. 대표적 기능인 음성인식은 화자종속이고 그 인식률은 99 %의 높은 인식률을 얻었다. 그 외에도 음성합성, 음악합성, 음성녹음 및 재생 등의 기능구현을 하였다. 음성인식을 가진 Intelligent Toy 계열의 시스템과 같은 잡음 환경 하에서 인식률을 비교해 볼 때 그 결과가 우수함을 확인하였다.

1. 서론

인간이 아닌 기계가 인간과 상호작용을 할 수 있게 하는 연구는 음성인식이라는 분야로서 많이 연구되어 왔다. 음성인식의 여러 분야 중에 현재 지능형 장난감의 개발은 내장형 시스템으로 국내외에서 일부 진행되어 상품화가 되었다.

이 연구 개발은 현재 다른 연구 개발의 중요핵심 부분으로 진행되는 지능형 장난감의 내장형 시스템 개발의 핵심부분이 된다. 내장형 시스템의 하드웨어와 소프트웨어 중에 소프트웨어 부분이 되는 것이다. 지능형 장난감에 내장될 소프트웨어는 음성과 각종 감지기의 상호작용을 통해 춤, 음악, 음성저장 등의 기능을 제공한다. 이를 통해 사용자와의 교감과 오락을 즐길 수 있는 방식의 지능형 장난감에 내장될 음성인식을 필두로 한 소프트웨어 개발이 이 연구 개발의 목표가 된다. 지능형 장난감의 하나는 Character Toy이고 다른 하나는 Character Toy를 내장할 UFO Toy이다. 소프트웨어 가능의 전부는 Character Toy 시스템에 내장된다.

관련 연구는 음성인식과 내장형 시스템에 대해 기술하였고 지능형 장난감의 소프트웨어 개발을 위한 시스템의 설계에 대한 내용은 상태 전이도, 소프트웨어 구조도, I/O 매핑으로 설명하기로 한다. 그에 따라 지능형 장난감의 소프트웨어 가능의 구현 및 통합 테스트한 내용을 기술하고 결론 및 향후 과제에 대한 제시로서 이 연구 논문의 마무리를 한다.

2. 관련연구

2.1 음성인식 기술의 종류 및 특징

음성인식은 응용분야와 사용기술에 따라 여러 종류가 있다. 응용분야에 따른 대표적인 종류로서 특정화자만을 인식하는 화자종속 음성인식, 여러 사람(일반인)의 말을 인식하는 화자독립 음성인식을 들 수 있다. 화자종속 음성인식은 화자독립 음성인식에 비해 인식률이 높아 실용화하기에 유리하다.

사용기술에 따라 음의 상태가 한 상태에서 다른 상태로 바뀌는 것을 천이확률로 표현한 HMM과 입력과 출력의 비선형관계를 함수로 표현한 특성을 지닌 신경망이론이 있는데 서로의 단점을 보완하기 위해서 둘을 결합한 모델이 많이 이용된다[1][2][3][5].

2.2 채택된 음성인식 기술

이 연구개발에 사용된 음성인식 기술은 신경망 이론이다. 신경망 학습의 종류에는 여러 가지가 있는데 가장 일반적인 방법으로는 감독 학습(Supervised learning)과 무감독 학습(Unsupervised learning)이 있다. 이 연구 개발에서는 두 종류가 모두 사용된다. 음성인식 기술에 신경망을 사용함으로써 생기는 이점은 다음과 같다.

1) 화자독립 음성인식 : 화자들의 공통되는 부분을 바탕으로 훈련(Training)함으로써 이 신경망은 모든 화자들의 공통적인 음성 특징을 추출할 수 있다. 그렇게 함으로써 특정 화자들의 악센트(Accent), 방언(Dialect), 피치(Pitch)가 무시될 수 있다. 결과적으로 목적물은 사용자 특성에 맞는 훈련을 필요로 하지 않는다.
2) 인식 접근방식의 탁월함 : 신경망의 수행은 화자독립 음성인식을 수행하기 위해 접근하는 다른 방식보다도 뛰어나다.
3) 저 비용 : 신경망의 연결 가중치는 음성신호의 본질적인 특성의 최소한의 표현이다. 즉, 인식을 수행하기 위해 필요한 메모리의 량은 작다. 더욱이 신경망은 특별히 제작된 RISC 프로세서를 사용해서 구현되어지므로 값비싼 DSP로의 구현보다는 훨씬 효율적이다[2][6].

2.3 내장형 시스템

이번 연구 개발은 앞서 서론에서 설명한 바와 같이 시스템

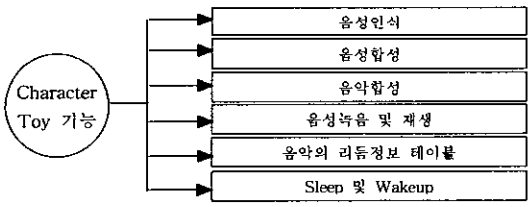
통합 시에는 하나의 내장형 시스템을 개발하는 것이 최종 목적이므로 그에 따른 내장형 시스템을 현재의 추세에 비추어서 설명하기로 한다. 일부 산업계에서만 쓰이던 내장형 시스템이 현재 Post-PC 시장의 폭발적인 수요로 인해 그에 대한 관심이 높아지고 있다. 기존 산업계에서는 몇몇 주류 Real-Time Operating System 업체가 내장형 시스템의 운영체제를 독점하다시피 하였으나, Post-PC 시장에는 아직 확실한 대안이 없이 여러 업체가 난립하고 있는 실정이다.

3. 연구 개발의 내용

3.1 소프트웨어의 기능

이 연구 개발에 쓰인 음성인식 기술은 화자종속 음성인식 기술이다. 사용자가 Character Toy 시스템에 인식할 단어를 말하면 그 단어에 대한 인식결과로써 응답을 하게 된다. 응답 및 반응의 종류에는 음성, 음악, 행동이 있다. 주변의 잡음에 대한 제한은 소음측정 단위(dB)로서 인식할 수 있는 환경의 최저 조건을 제시하기로 한다.

Character Toy의 기능 중 음성합성과 음악연주 기능은 전처리 과정으로 사운드 데이터를 RSC-364에서 처리할 수 있는 규격으로 만들기 위해서 샘플링 및 필터링을 거쳐야 한다. 그에 따른 도구는 상용 사운드 에디터 프로그램과 Development Kit에서 제공되는 소프트웨어인 Quick Synthesis Tool을 이용한다. 메모리의 공간을 절약하고 음질의 수준이 고려되어야 하므로 음성압축을 하는데 4 비트, 3 비트, 2 비트의 압축이 가능하다. Sleep 및 Wakeup의 기능은 사용자와 상호작용을 할 수 있고 또 하나는 전원절약을 할 수 있다는 것이다. 배터리의 특성상 동작 시간이 제한되어 있으므로 필요 없는 전원의 낭비를 줄일 수 있다.



[그림 1] 소프트웨어의 기능에 대한 개요도

3.2 소프트웨어 개발 칩 및 도구

3.2.1 RSC-364

개발할 시스템에 내장될 기능을 처리할 음성인식 칩은 미국의 Sensory사에서 개발된 RSC-364라는 칩이다. 일반 DSP 칩으로 음성인식 등 여러 가지 기능을 구현하게 되면 복잡성, 시간상의 제한, 호환성의 결핍 등의 문제가 발생된다. 이 문제로 인해 음성인식 등 여러 가지 기능을 갖춘 음성인식 칩을 이용하여야 되는데 그 기능을 모두 갖춘 칩으로는 RSC-364가 유일하다. 이 칩은 4MIPS 8비트 프로세서이고 메모리는 2.5 Kbyte RAM과 64Kbyte 내부 ROM이 있다. 칩 내부에 사운드 출력을 위한 A/D 컨버터와 D/A 컨버터가 내장되어 있어 디지털 및 아날로그 신호변환 처리가 될 수 있고 사운드 입출력의 증폭을 위한 전치 증폭기(Preamplifier)가 있다. 외부 메모리 버스로는 16비트 주소 버스와 8비트 데이터 버스를 사용한다[6].

3.2.2 RSC-364 Development Kit

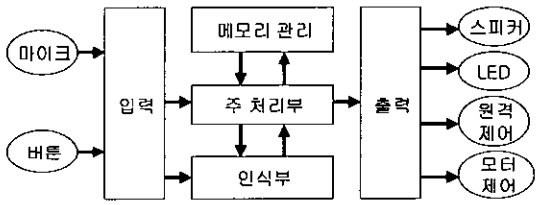
개발도구인 RSC-364 Development Kit의 구성은 총 3가지의 모듈로 되어 있다. RSC-364칩이 내장되어 있는 마더보드, 실행 데이터의 저장을 위한 메모리 모듈, I/O 모듈이 있다.

지능형 장난감의 소프트웨어의 각 모듈의 구현은 PC상의 오프라인에서 8051계열의 Assembly어로 수행한다. 컴파일된 실행파일은 시리얼 포트를 통해서 Development Kit에 저장하여 실행한다. 프로그램이 내부적으로 수행되는 과정을 PC상에서 오프라인으로 파악하기 위해서는 RSC-364의 전용 시뮬레이터를 사용한다[6].

4. 소프트웨어 개발 위한 시스템 설계

4.1 Toy 시스템의 구조도

Toy 시스템의 전체 기능 중 하나는 Character Toy 시스템이고 다른 하나는 UFO Toy 시스템이다. UFO Toy 시스템은 하드웨어로 개발되기 때문에 자세한 설명은 생략한다. Toy 시스템의 전체 구조도는 [그림 2]와 같다. 주 처리부는 입력부와 인식부의 결과에서 들어온 결과를 처리하여 출력부로 결과를 출력한다. 메모리 관리부는 플래시(Flash) 메모리 드라이버와 섹터관리를 하는 부분이다. RSC-364는 16Bit의 주소 버스를 가지고 있어 기본적으로는 64Kbyte의 메모리를 사용할 수 있으나 Toy 시스템은 4Mbit 메모리가 필요하다. 메모리 관리부에서는 IO Mapping Table을 제어하여 큰 메모리를 인식하도록 한다. 인식부는 음성 인식을 하고 인식의 결과에 따라서 재인식과 인식결과를 주 처리부로 전달한다. 출력부에서 원격제어는 RF(Radio Frequency)통신을 통해 UFO를 제어하고, 모터 제어부에서는 Toy의 모터를 제어한다.



[그림 2] 개발할 시스템의 소프트웨어 구조도

4.2 Toy 시스템의 입/출력 제어

RSC-364 칩의 장점으로는 General purpose IO를 지원하는데 있다. 본 연구에서는 GPIO를 이용하여 원격 제어회로와 모터 구동회로를 제어한다. RSC-364는 16개의 GPIO를 지원하고 있다. 하지만 Toy 시스템의 IO의 수가 16개 보다 많기 때문에 IO핀을 Module화(폴링정책)하여 사용한다. [표 1]과 [표 2]에서 IO 핀들의 할당 영역을 보여준다.

[표 1] RSC-364 외부I/O에 사용된 입력핀

P0.0	P0.1	P0.2	P0.3	P0.4	P0.5	P0.6	P0.7
interrupt	입력모드	입력 버튼 영역			확장 주소 영역		

[표 2] RSC-364의 외부 I/O에 사용된 출력 핀

P1.0	P1.1	P1.2	P1.3	P1.4	P1.5	P1.7	P1.6
UFO 제어		LED 제어			모터 제어		Stop

5. 구현 및 통합시험

5.1 음성합성

음성합성과 음악연주에 대하여 공통으로 적용한 기능이다. 음성 데이터는 일반적인 PCM(Pulse Code Modulation)코딩을 사용하지 않고, ADPCM 코딩 방법을 사용한다. Toy 시스템에 많은 음성 정보를 저장하기 위해서 ADPCM 코딩은 필수적이다.

그리고 음성 데이터는 4bit로 코딩을 하였으며, 음악 데이터는 3bit로 압축하였다. 합성된 사운드 데이터의 음질분석을 한 결과 음악의 경우에는 사운드를 최소 3비트로 압축했을 때에 들는데 지장이 없었고 음성의 경우에는 최소 4비트로 압축하여야 제대로 된 음질을 들을 수 있었다.
이때 사운드 파일은 8000Hz, 16비트, 모노로 데이터 포맷이고, Sensory사에서 제공하는 "Quick Synthesis Tool"을 이용해서 잡음제거를 위한 IIR 필터링을 시키고 음성 압축을 한다. 잡음제거의 또 하나의 방법으로 사운드의 파형(Waveform)에서 진폭(Amplitude)을 조정하는 기술을 사용하였다[4][7].

5.2 음성인식

인식할 단어의 수는 총 5개이다. 인식 수행 시 저해 요건 중에서 가장 크게 영향을 끼치는 잡음에 대한 강인성을 테스트 해 보았다. 인식단어에 대한 인식률은 각 기준에 따라 테스트를 하였다[표 3]. 그 기준은 특정인의 경우는 1명에 대한 횟수 100번 기준으로 인식률을 산출하였고 일반인의 경우는 20명에 대한 사람 수를 기준으로 테스트한 결과를 인식률로 산출하였다. 테스트를 한 결과 화자종속인 특정인에 대한 인식률은 잡음이 거의 없는 경우는 99%의 인식률을 보였고 잡음이 80dB인 환경에서는 90%의 인식률을 보였다. 이 연구개발에 쓰인 음성인식 기술이 화자종속이지만 일반인에게도 인식률 테스트를 해 보았다. 그 결과 정확한 발음의 경우는 90%의 인식률을 보이지만 비 정확한 발음이나 잡음이 섞이는 경우에는 14%의 저조한 인식률을 보였다.

음성인식에서 가장 핵심 기술은 Weight Table을 생성하는 부분이다. 이 개발은 특정화자의 인식할 단어에 대한 훈련 과정에서 생성된 Weight Table을 추출해서 인식 모듈에서 자유자재로 쓸 수 있게 하는 것이 관건이었다. 추출된 Weight Table은 화자종속인식 모듈에 이용하였다.

[표 3] 음성인식 단어에 대한 각 기준별 인식률

인식인의 구분	특정인		일반인 (발음정확)	일반인 (유사단어 및 보통이하 발음)
	보통의 잡음환경	잡음이 80dB인경우		
인식률	99%	90%	90%	14%

5.3 음성녹음 및 재생

음성녹음 및 재생 모듈은 일반음성 녹음 및 재생 모듈과 러브메시지 녹음 및 재생 모듈의 두 부분으로 나뉘어 진다. 각 모듈의 기능은 동일하다. 다만 메모리의 통합 및 저장될 데이터의 위치가 달라지게 된다. 최대 녹음시간은 7초로 하였다. 녹음될 데이터는 4비트로 압축하였고 초당 4067 byte 의 최대 데이터 압축률을 가진다.

5.4 리듬정보 테이블 생성

Character Toy 시스템의 음악연주에 맞추어서 Character Toy의 모터가 구동하게 만들었다. 음악에 따라 다른 리듬 정보 테이블 생성을 오프라인에서 수동으로 미리 만든 다음에 실제 동작 시에는 만들어진 리듬정보 테이블에 의해서 모터 구동신호를 전송하게 하였다. 리듬정보에는 박자, 음의 세기, 빠르기 등이 있는데 테이블의 구성 기준은 시간의 간격에 두었다. 모터 구동 신호를 보냈다가 끊었다가 하여 춤동작을 수행하게 하는 것이다.

5.5 디바이스 드라이버 및 I/O 모듈

데이터 저장을 위해서 플래시메모리를 사용하였는데 그 종류에 따라 저장하고 호출하는 방식이 틀리다. 그에 맞게 디바이스 드라이버를 조정하고 수정하는 작업을 하였고 Data Bank Switching 등 부가적으로 각 기능별 모듈도 그에 따라 수정하였다.

이 장에서 기술한 음성합성, 음성인식, 음성녹음 및 재생, Sleep 및 Wakeup, 리듬정보 테이블 생성의 핵심기능의 모듈을 전체 시스템으로 통합하여 테스트해 보았다. 현재는 Prototype으로 완성되어 있지만 차후 제품화된다면 좋은 제품이 되리라 생각한다.

6. 결론 및 연구과제

기존의 음성인식만의 기능을 가진 지능형 장난감에 비하면 음성녹음 및 재생, 음악연주의 다양한 기능을 가진 이 연구개발의 성과는 크다고 할 수 있다. 각각의 기능에 대한 세부적인 모듈의 구현은 많은 시간이 요구되지 않았으나 전체 모듈의 통합에는 메모리의 특성과 그에 따른 하드웨어적인 부가작업이 필요 되어 많은 시간이 요구되었다.

이 연구개발의 주안점은 지능형 장난감에 내장될 핵심기능의 구현 및 통합, 추가적인 모듈의 구현, 메모리의 효율적인 활용이었다. 순수 자체 연구 개발된 추가 모듈 중에서 리듬정보 테이블의 생성은 수동으로가 아닌 자동으로 생성하는 모듈을 구현하는 것이 앞으로 해결해야 할 과제이다.

[참고 문헌]

[1] A. Acero, and R. M. Stern., "Environmental Robustness in Automatic Speech Recognition," ICASSP-90, P.849-952, 1990.

[2] A. F. Murray, "Applications of Neural Networks," Kluwer Academy Publishers, P.1-28, 1995.

[3] J. G. Wilpon, P. Mikkilineni, D. B. Roe, and S. Gokcen, "Speech Recognition: From the Laboratory to the Real World," AT&T Tech., Vol.69, No.5, P.14-24, 1990.

[4] R. J. McAulay, and T. F. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," IEEE Trans. Acoust. Speech Signal Proc. Vol.34, No.4, P.744-754, 1986.

[5] R. P. Ramachandran, and R. Mammone, "Modern Methods of Speech Processing," Kluwer Acdemic Publishers, P.121-299, 1996.

[6] Sensory Inc. , "RSC-364 Development Kit Manual with Sensory Speech 5.0 Technology," Chapter 3, P.1-48, 1999.

[7] S. R. Quackenbush, T. P. Barnwell, and M. A. Clements," Objective Measures of Speech Quality," Prentice Hall, P.750-753, 1998.