

## 저가의 단 문장 음성 인식회로 설계

최지혁, 홍광석

성균관대학교 정보통신공학부 휴먼컴퓨터연구실

Tel : 031-290-7196 / HP : 016-368-6169

### Low Cost Circuit Design for a Sentence Speech Recognition

Ji-Hyok Choi, Kwang-Seok Hong

HCI Lab., School of Information & Communication Engineering, Sungkyunkwan Univ.

sbncap@mail.skku.ac.kr, kshong@skku.ac.kr

#### Abstract

In this paper, we present a low cost circuit design for a sentence speech recognition. The basic circuit of the designed sentence speech recognizer is composed of resistor, capacitance, OP Amp, counter and logic gates.

Through a sentence recognition experiment, we can find the effectiveness of the designed sentence recognition circuit.

#### I. 서론

컴퓨터 하드웨어 기술의 급속한 진보와 음성처리 기술의 발전으로 인하여 음성인식의 실용화가 실질적인 문제로서 관심이 증대되고 있다. 이러한 관심이 증대되면서, 음성인식에 관한 연구는 실용화에 초점이 모아지면서 최근 몇 년간 눈부시게 발전하여 일부 데스크에서 상업용 시스템이 구현되고 있는 실정이다.[1,2]

상업용 시스템은 컴퓨터기반의 고성능 프로세서를 이용하여 숫자음성인식 시스템, 화자 인식 시스템, 전문가 시스템이 구현되어 사용되어지고 있고, 지능을 가진 십자 말게임, 스무고개 게임, 끝말 이어가기 게임 등이 연구되어 지고 있다. 이러한 시스템은 안정성과 인식률의 성능이 높아야 하므로 상당한 고가의 시스템

으로 구현되어져 있다. 그리고 중 고가의 완구용이나 학습용 제품으로는 DSP나 PDA를 이용한 어학 학습기, 명령어 실행기가 주류를 이루고 있는 실정이다.[3-6]

이에 반하여 저가의 완구형 시스템에 적용할만한 소형의 시스템은 시장성에 비하여 아직 개발이 미비한 상태이다. 저가의 완구형 시스템의 경우 대용량 인식기와 같은 가변어휘형 무제한 인식과 같은 고성능보다는 특정한 단어나 한 문장에 대한 검출만을 원하는 경우가 대부분이다. 이러한 소형 시스템은 대형 시스템에서 필요한 메모리와 고속 연산 수행능력 필요치 않다.

본 논문에서는 특정 문장 또는 단어에 대하여 음성의 유무 판단회로, 잠음 판단회로, 소구간 및 대구간 에너지 검출 및 카운터, 그리고 인식 판단 회로를 OP-Amp를 이용하여 설계하였다. 또한 이를 이용하여 저가의 제품에 적용할 수 있는 단 문장 음성인식 회로를 설계하였다.

#### II. 음성인식 시스템

##### 2.1 음성인식 방법

지금까지 개발된 음성인식 알고리즘은 크게 세 가지로 나누어 볼 수 있다. 첫째로는 동적 프로그래밍법을 이용하여 기준이 되는 음성과 입력 음성을 비교하여

최소거리 법칙에 의하여 입력 음성을 인식하는 DTW(Dynamic Time Warping) 방법이다. 이 DTW 알고리즘을 이용하여 격리 단어뿐만 아니라 연결단어 까지도 인식하기 위하여 OSDP(One-Stage Dynamic Programming)방법이나 LBDTW(Level-Building Dynamic Time Warping)방법등이 제안되었다. DTW를 이용하는 음성인식 방법은 높은 인식률을 얻을 수 있다는 장점은 있으나 많은 수의 기준 패턴을 구성하기 위해서는 많은 시간과 노력이 필요하며 인식과정에 걸리는 시간이 길기 때문에 100여 단어 정도의 어휘를 화자 종속으로 인식하는 경우에 적합하나 대용량 단어의 인식에는 적합하지 못하다는 단점이 있다.

두 번째로 신경망(Neural Network)을 이용하여 음성을 인식하는 방법이 최근에 제안되어 주목을 받고 있다. 신경망이 갖는 장점은 병렬처리가 가능하므로 계산속도를 높일 수 있다는 점과 신경망의 성능을 입력 데이터가 주어짐에 따라서 점진적으로 향상시킬 수 있도록 가중치를 학습시킬 수 있다는 것이다. 이러한 이유로 신경망을 이용하여 음성을 인식하는 연구가 많이 진행되어지고 있다.

세 번째로는 HMM(Hiden Markov Model)을 이용하여 음성을 확률적으로 모델링 하는 방법이 있다. HMM을 이용한 음성인식 시스템의 성능을 향상시키기 위하여 음성에 관한 음소의 지속시간, 음운의 탈락, 변화와 HMM과정에 사용되는 벡터 양자화 과정의 왜곡을 없애거나 줄이기 위한 방법 등 여러 가지 방법이 제안되고 연구되고 있다.

## 2.2 음성인식 회로

현재의 음성인식은 대부분 2.1절에 소개한 세 가지 알고리즘을 가지고 인식시스템을 구현하게 되는데 HMM을 이용한 가변어휘 인식이 주류를 이루고 있고, HMM자체가 확률 모델을 기반으로 하는 알고리즘이므로 많은 연산과 메모리 량이 필요하기 때문에 임베디드 시스템보다는 PC를 기반으로 한 시스템에서 높은 성능을 보이고 있다.

임베디드 제품으로 출시되는 음성인식 보드의 몇 가지 제품을 살펴보면 40MIPS이상의 성능을 가지는 DSP를 이용하여 50단어 내외의 화자 독립형 음성인식기[7]가 출시가 되고 있다. 이러한 제품의 경우 보드와 시스템 알고리즘 구성을 포함한 제품의 가격이 저가의 완구형 제품에 적용하기에는 높은 가격 대를 형성하고 있어 PDA나 중 고가의 장비에 적용을 목적으로 제품이 출시되고 있다. 10단어 내외의 인식 제품이 실용화 되기도 했다.[8]

이상의 제품들은 오디오, 핸드프리, PDA 시스템, 음

성 도어록의 제품에 사용되어 지고있으나 저가의 완구형 제품에 적용할만한 저 가격의 시스템은 구현이 미비한 상태이다.

## III. 저가의 음성인식 회로 설계

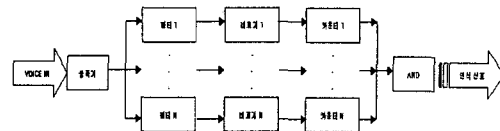


그림 1 음성인식 회로 계통도

본 논문에서 제안한 저가의 음성인식기는 OP-Amp를 이용하여 음성의 유무, 잡음판단, 소구간 및 대구간 에너지 등의 특징을 파악하여 인식 판단을 하는 시스템이다. 그림 1은 입력 음성 신호에 대한 필터뱅크를 구성한 시스템 계통도이다.

### 3.1 주파수 검출

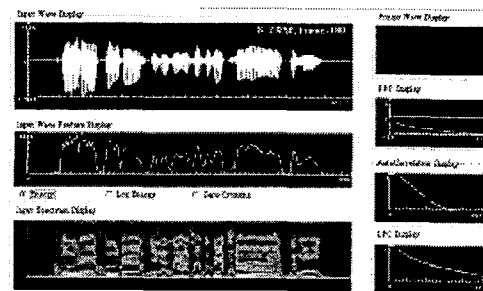


그림 2 '거울아~ 거울아~ 이 세상에서 누가 제일 예쁘니' 파형과 스펙트럼

본 시스템은 고가의 시스템에서처럼 주파수 분석을 자체에서 할 수가 없으므로 인식하고자 하는 음성의 주파수를 미리 분석하여야 시스템을 설계할 수가 있다. 그림 2는 STFT를 이용하여 원음 신호를 분석하였다. 해밍윈도우를 사용하고 256개의 데이터를 한 Frame으로 하여 50%중첩을 한 스펙트럼이다. 그림 2를 통하여 OP-Amp의 필터 대역을 선택하게 된다.

### 3.2 대구간 에너지 검출 및 카운터

그림 2를 통하여 음성의 검출 분장의 특징을 파악한 후 그림 1의 구조로 필터뱅크를 구성하게 된다. 그림 3은 마이크로 들어오는 음성신호를 일정 크기의 비로 증폭을 하고 증폭된 신호를 이용하여 필터 회로를 거

치게 된다. 그림 3의 필터회로는 대구간 에너지를 검출하는 필터 회로이다. 필터를 거친 신호는 비교기를 통해 디지털 값으로 변환하고 카운터를 이용하여 문턱 값과 비교하여 최종 신호를 출력하게 된다. 대구간 에너지 값은 그림 1에서 보이는 단어와 단어 사이의 에너지 차를 판단하는 동작을 하게 된다.

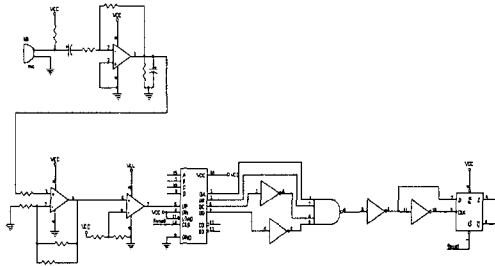


그림 3 대구간 에너지 검출 및 카운터 회로

### 3.3 소구간 에너지 검출 및 카운터

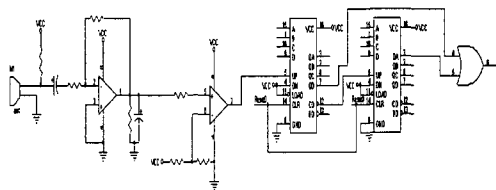


그림 4 소구간 에너지 검출 및 카운터 회로

소구간 에너지 검출은 3.2에서 설명한 대구간 에너지 검출회로와 구성은 거의 비슷하다. 다른 점은 비교기의 임계값을 작게 주어 소구간 에너지를 검출하여 음절을 구분하게 된다. 대구간 에너지 카운터 값보다 많은 카운터 값을 가지게 된다.

### 3.4 음성 유무 및 잡음 판단

일반 음성인식 시스템에서도 중요한 문제로 인식되고 있는 끝점검출과 유사한 기능을 하는 회로이다. 저가의 완구용 시스템일 경우 따로 ADC를 이용한 영교차율 값을 계산하는 회로를 첨가하게되면 다른 회로보다 높은 비용을 들여야한다.

본 논문에서는 마이크 입력을 증폭한 후 실제 입력이 일정시간 없을 경우 대구간 에너지 카운터회로와 소구간 에너지 카운터 회로의 초기화 신호를 보내는 방법으로 잡음 입력시의 리셋트 신호와 음성 검출 신호로 사용하였다.

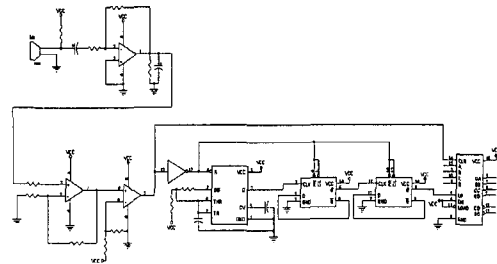


그림 5 음성 유무 및 잡음 판단 회로

## IV. 실험 결과

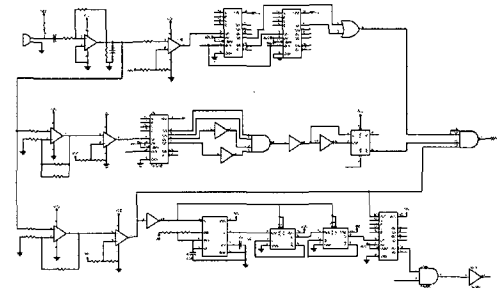


그림 6 단문장 인식회로

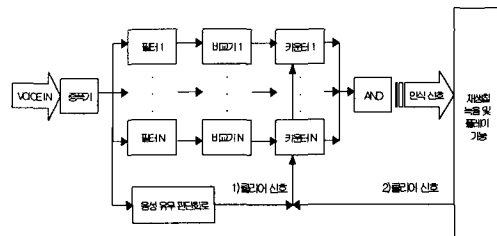


그림 7 단 문장 인식회로의 전체 계통도

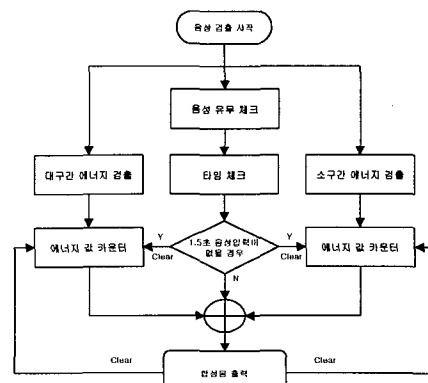


그림 8 회로의 순서도

그림 6은 앞에서 설명한 그림 3, 4, 5를 통합하여 음성이 들어왔을 경우 최종 결과를 출력하는 회로 구성하였다. 그림 6의 회로를 이용하여 그림 7과 같은 응용회로를 구성하여 실험을 하였다. 응용 실험의 순서도는 그림 8과 같다. 실험을 위한 문장은 '거울아~ 거울아~ 이 세상에서 누가 제일 예쁘니?'라는 문장을 22세 미만의 남녀 30명을 대상으로 실험적으로 얻은 초기 치를 적용하여 인식 회로를 설계하였다.

인식 실험은 초기 치를 얻을 때 사용하지 않은 남녀 15명이 10회씩 발생하였고 실험방법은 두 가지로 하였는데, 일반적으로 '거울아 거울아 이 세상에서 누가 제일 예쁘니?' 라고 억양을 주지 않고 발생하기 때문에 발생연습을 하지 않고 실험하는 경우(실험 1)와 '거울아~ 거울아~ 이 세상에서 누가 제일 예쁘니?' 라는 억양을 주는 문장음성의 발생을 3회 연습하고(실험 2) 실험을 한 결과이다. 인식되지 않은 것은 회로가 아무 반응이 없이 입력음성을 거절한 경우이다.

표 1. 발생연습 전의 인식실험 결과

	인식률
남	32.67%
여	25.33%

표 2. 발생연습 후의 인식실험 결과

	인식률
남	57.33%
여	59.33%

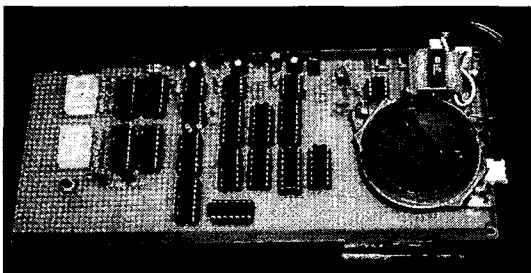


그림 9 제작한 인식실험 회로

## V. 결론

본 연구의 목적은 낮은 가격으로 대량 생산이 가능한 완구제품의 시스템에 적용이 가능한 한 문장 음성 인식 회로의 설계이다.

본 논문에서 제안한 시스템은 회로 기본소자와 OP-Amp를 이용하여 저가의 회로를 구성할 수가 있다. 또 이러한 회로의 장점은 저가의 완구형 제품에서 일어날 수 있는 소비자들의 부주의로 인하여 DSP나 마이크로 프로세서를 이용한 제품들보다 고장에 대하여 부담이 적은 장점이 있다.

전원의 제한(4.5 볼트)으로 인하여 BPF를 구성할 때 일반 필터 설계이론에 따르지 않아서 실험적으로 몇 가지를 구성하였으나 인식률에 큰 향상을 보이지 않아서 회로에서 제거하였는데 좀더 높은 인식률이 필요한 응용에서는 추가해야 할 것으로 생각된다.

앞으로 인식률의 개선과 회로의 집적화를 통해 다양한 한 문장을 인식할 수 있는 IC로 구성한다면 저가의 완구형 제품에 다양하게 적용할 수 있는 유용한 회로라고 볼 수 있다.

### <감사의 글>

본 연구는 (주)디지텍의 연구비 지원으로 이루어 졌습니다.

## 참고문헌

- [1] 정한열, "음성인식 연구의 국내외 현황과 전망," 제 15회 음성통신 및 신호처리 워크샵 논문집, pp. 23-30, 1998
- [2] 김순협, "음성인식의 현황과 최근 연구 동향," 2000년도 한국 음향학회 학술발표대회 논문집, Vol. 19, No. 2(s), 2000
- [3] 박상훈, "PAD상에서 음성명령어를 구현하기 위한 음성인식기의 설계," 2001년도 한국 음향학회 학술발표대회 논문집 Vol. 20, No. 2(s), 2001
- [4] 윤재선 "한국어 음성인식 Dictation System의 ," 성균관대학교 박사학위 논문, 2001
- [5] 김동주, 이용주, 김동환, 윤재선, 홍광석, "음성인식 및 합성을 이용한 십자말 게임의 구현", 한국음향학회 추계학술발표, 20권 2호, pp29-32, 2001
- [6] 윤재선, 김동환, 홍광석, "음성인식 끝말 이어가기 게임의 구현", 한국 신호처리 · 시스템학회, 1권 2호, pp177-180, 2000
- [7] www.slwordl.co.kr
- [8] www.3ms.co.kr
- [9] L. Rabiner and B.H. Juang, "Fundamentals of speech recognition", Prentice Hall, 1993.