

동굴관광용 고층수직이동 승강기의 긴급 음성구동 제어

최병섭 · 강태현 · 윤여훈 · 장훈규 · 소대화*

Voice Recognition Sensor Driven Elevator for High-rise Vertical Shift

Choi Byong-Seob, Kang Tae-Hyun, Yun Yeo-Hoon, Jang Hoon-Gyou, Soh Dea-Wha

Abstract : Recently, it is one of very interest technology of Human Computer Interaction(HCI). Nowadays, it is easy to find out that, for example, inside SF movies people has talking to computer. However, there are difference between CPU language and ours. So, we focus on connecting to CPU. For 30 years many scientists experienced in that technology. But it is really difficult. Our project goal is making that CPU could understand human voice. First of all the signal through a voice sensor will move to BCD (binary code). That elevator helps out people who wants to move up and down. This product's point is related with people's safety. Using a PWM for motor control by ATmega16, we choose a DC motor to drive it because of making a regular speed elevator. Furthermore, using a voice identification module the elevator driven by voice sensor could operate well up and down perfectly from 1st to 10th floor by PWM control with ATmega16. And, it will be clearly useful for high-rise vertical shift with voice recognition sensor driven.

Key words : voice recognition sensor, elevator, human computer interaction, CPU, PWM, speaker recognition

I. 서론

1. 추진배경

최근 Human Computer Interaction(HCI)에 관한 관심이 증대하면서, SF영화에서 볼 수 있었던 사람의 음성으로 컴퓨터와 대화를 하는 것에서 동기를 얻어서 음성으로 기계를 동작시키는 것에 관심을 두고 프로젝트를 시작하게 되었다. 그러나 컴퓨터와 사람의 사용 언어가 다르기 때문에 이를 연결시켜주는 방법에 먼저 초점이 맞춰지게 되었다. 즉, 음성인식 기술은 컴퓨터가 사람의 귀와 같은 기능을 갖게 하는 인공지능의 첨단 기술이다. 세계적으로 30년 이상 연구되었으나, 실용사례를 찾아보기 힘든 어려운 기술이며, 따라서 기술적 결과가 적용되면 폭발적인 시장성을 확보할 수 있을 것이다.

음성이 다른 인터페이스보다 뛰어난 점은 첫째, 편의성이다. 인간이 음성으로 기계와 대화 하듯 의사소통이 가능하다는 장점이 있다. 둘째 병렬성이다. 음성으로 대화하며 키보드 치는 등의 동시적 처리가 가능하며, 정보를 사용하거나 입력하는데 자유로이 움직이면서 정보의 입출력이 가능하다는 장점이 있다. 셋째 자료입력의 고속화와 원거리 입력이 가능하다. 음성은 글 쓰는 것보다 4배정도 더 빠르고, 전화선이나 무선통신을 통한 원거리입력이 가능하기 때문이다.

따라서 컴퓨터가 인간의 음성으로 내려진 명령을 수행하는 것을 구현하는데 목적을 둔다. 구현을 위한 구체적인 대상은 실험용으로 직접 제작한 시작품 엘리베이터를 대상으로 삼았다.

음성인식 감지 동작의 실험결과, 음성신호 특징의 규정에 따라 다소 차이가 있으나 지정한

* 명지대학교 교수

의 명령에 따라서 음성인식 감응으로 승강구간의 층 간 운행이 자연스럽게 작동됨을 확인하였으며, 운행 중 긴급 상황발생 시 육체적 동작에 따른 긴급 명령 전달체계의 지연시간 발생을 음성신호 전달방식으로 대응함으로써 이를 충분히 신속하게 단축/제거하여 승강운행 중 돌발적 응급/긴급 상황을 보다 효과적으로 대처할 수 있음이 고층수직승강기에 유익함으로 확인되었다.

II. 본론

음성인식을 이용한 엘리베이터 제작을 위하여 먼저 시스템 구성을 분류하였다. 화자가 음성으로 명령하면 음성을 인식할 수 있는 음성인식부, 인식한 음성신호를 CPU(Atmega16)에 전송한다. 그리고 CPU를 통해서 LCD를 통한 현재의 위치와 그리고 모터를 통한 승강기의 층수를 나타내는 구동부를 제어하고, 각각의 음성인식부, 제어부, 구동부를 구성한다.

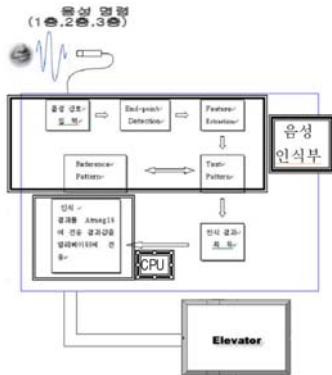


그림1. 음성인식 엘리베이터의 구성도

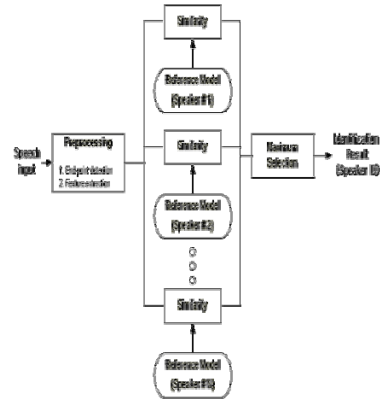


그림2. 화자인식시스템 흐름도

1. 음성 인식부

음성인식센서를 이용하여 음성인식을 구현하였다. 우리가 사용하는 음성인식센서의 음성인식이란 어떤 것인가에 대해서 간단히 설명하고 음성인식방법 중에서 우리 조가 사용하는 JT-2000의 특징을 설명하고 J-2000을 통한 음성센서 신호를 해석한다.

1.1 화자인식방법

화자인식은 식별되어야 할 화자들 간의 발음으로부터 데이터를 모으는 작업인 훈련과정과 임의의 발음을 식별하는 판별과정으로 이루어진다. 화자인식은 크게 두 가지로 나누어지는데 closed-set problem과 open-set problem 이다. Closed-set problem 이란 N 명의 알려진 화자 중에서 어떤 사람인지를 식별해 내는 것이다. 이 경우 N 이 커지면 커질수록 식별하는 것은 어려운 문제가 될 것이다. 이와 달리 open-set problem 은 식별해 내고자 하는 화자가 N명의 알려진 화자의 그룹에 속하는가 하는 것을 판단해야 하는 문제이다. 이 두 문제에 대해서 전자는 화자식별 후자는 화자검증이라고도 부른다. 화자 검증시스템은 알려진 화자의 그룹에 속 하는가 아닌가에 대해 이진결정을 내리기만 하면 되지만, 녹음

한 음성을 이용한 사기를 막기 위해 음성인식 또한 동시에 요구되는 경우가 많다. 이러한 화자 식별 및 화자검증 기술은 개인의 음성 특징이 유일하다는 사실을 근거로 하고 있으며 최근의 인터넷 기술의 발전과 더불어 보안을 위한 인증 방법으로 각광을 받고 있다. 화자인식시스템은 발성의 종류에 따라 문맥종속 및 문맥독립화자인식으로 나눌 수 있는데, 문맥독립화자인식의 경우 보안성이 높아 이에 관해 많은 연구가 진행 중에 있다.

문맥독립화자인식방법은 식별과정이 어떠한 text에도 동작하는 인식 방법을 말한다. 문맥종속화자인식방법과는 혼련, 테스트 방법이 다르다. 문맥종속화자인식방법에 있어서는 성능향상을 위해서 발음할 내용이 조절될 수 있다.

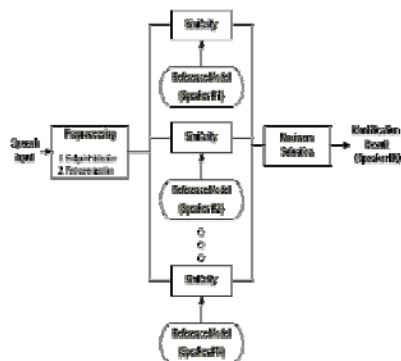


그림2. 화자인식시스템 흐름도

화자인식 방법에는 Template matching, Vector Quantization에 기반으로 하는 방법, HMM과 GMM(Gaussian Mixture Model)에 기반으로 하는 방법 등이 연구되고 있으며 이러한 접근방법들 중 화자특성변화의 표현에 있어서나 화자인식을 면에서 좋은 결과를 나타내고 있는 GMM에 의한 접근방법이 가장 유리한 것으로 알려져 있다.

1.2 DTW 기법을 이용한 화자인식 알고리즘

화자인식은 화자식별기술과 화자검증기술로 나눌 수 있다. 화자식별 기술은 고립 단어인식과 개념이 비슷하다. 고립단어인식은 발화된 음성과 가장 가까운 단어를 등록된 인식 대상 단어 중에서 찾아낸다.

마찬가지로 화자식별도 등록된 화자 중에서 가장 유사한 화자를 골라내는 것이다. 이 기술은 여러 명의 용의자 중 누가 범인인지 알아내는 데 사용될 수 있다.

화자 검증 기술은 핵심어인식처럼 승인과 거절과정을 거치게 된다. 이 과정은 기준패턴과 입력패턴을 서로 비교해 미리 정해 놓은 발생 확률 값을 넘으면 승인하고, 그렇지 않으면 거절하는 것이다. 이를 응용하면 음성 자물쇠로 이용할 수 있다.

화자인식 시스템을 어떤 형태로 구현할 것인가의 관점에서 보면, 문맥종속과 문맥독립으로 나눌 수 있다. (Table 1 참고) 문맥종속은 정해진 말, 즉 미리 정해 놓은 단어나 문장 등을 뜻한다. 문맥종속 시스템의 경우에는 그 특성 때문에 DTW(Dynamic Time Wrapping)Algorithm을 사용해 성능이 좋은 반면 다른 사람이 정해진 말을 엿듣고 흉내 낼 우려가 있다. 문맥독립은 미리 정한 말이 없이 아무 말이나 하는 것이다. 문맥독립 시스템의 경우 HMM(Hidden Markov Model)Algorithm을 많이 사용해 문맥종속 시스템의 단점을 감소시킬 수 있다. 참고로 미국의 경우 1,000명의 다른 사람(Impostor)이 시험한 결과 1명 이하의 사람을 잘못 승인하고, 100번 발생해 1번 이하의 잘못된 거절을 화자인식 시스템의 최소 규격으로 삼고 있다. 최근 발표되는 국제학술지에서 화자인식 기술의 수준이 앞에서 언급한 성능을 이미 넘어서고 있다.

<표 1> 문맥 종속형과 문맥 독립형의 비교

구 분	문맥 종속형 (Text dependent)	문맥 독립형 (Text Independent)
많이 상용하는 Algorithm	DTW	HMM
훈련에 필요한 DATA	짧은 발성	대규모 음성
장 점	성능이 좋음	아무 말이나 가능함
단 점	타인이 흉내 내기 쉬움	훈련 과정이 복잡

DTW 기법은 기준이 되는 음성신호의 패턴과 입력된 음성 신호간의 유사도를 동적 프로그래밍을 이용해 구하는 방법이다. 시간 축에서 차이를 보상하기 위한 방법이다. 예를 들어 길이가 M인 입력 음성 패턴을 $T=T(1), T(2), \dots, T(M)$ 길이가 N인 기준 패턴을 $R=R(1), R(2), \dots, R(N)$ 라고 하면 두 패턴 간의 유사도D는 다음과 같이 누적거리로 표현된다. 예를 들어 우리가 자주 사용하는 단어 중에는 “아주머니”라는 단어가 있다. 이 단어에 대한 동일 인물의 음성 파형을 그리면 그림3과 같다.

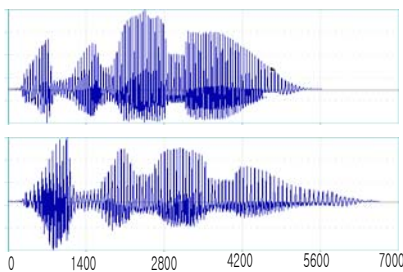


그림3. “아주머니”의 동일인 발성파형

이와 같이 동일인이 동일 단어를 발성했을 경우는 아래와 같은 문제점이 발생한다.

- 발성시마다 단어의 시간적 길이 변화
- 이를 참조패턴과 단순히 비교하면 오류/

인식불능 발생(시간축이 고르지 않음)

→ 시간축의 정규화 방법이 필요함.

이를 해결할 수 있는 방법이 DTW이다.

1.3 음성센서모듈(JT-2000)

음성 인식부를 구성하면서 소프트웨어적인 면과 하드웨어적인 면을 구현하는 것에 어려움이 너무 많고 시간이 부족하여 전원 테크에서 개발한 JT-2000 음성 인식 모듈을 구입하여 제작하였다. 구입한 JT-2000을 통해서 인식한 10개의 단어(1층부터 10층까지)를 이용하여 음성 인식한 단어를 BCD 코드 값으로 Atmega16에 전송한다.

2. 중앙처리장치-CPU

ATmega16 마이크로 컨트롤러는 ATMEL사에서 개발한 저가이면서 고속인 유사 RISC칩으로서, CMOS 형태로 이루어진 8비트의 마이크로 컨트롤러이다. AVR ATmega16의 특징과 ATmega16을 통한 전체적인 알고리즘을 구성하였다.

2.1 CPU를 통한 제어

음성센서를 통해서 인식된 신호를 BCD코드로 처리하여 제어한다. 엘리베이터는 사람이 쉽게 위, 아래층으로 이동할 수 있도록 도와주는 기계장치로써, 사람의 안전과 직결됨으로 신뢰성을 갖추는 것이 중요함은 당연한 일이다. 그러므로 부하에 따라 변하지 않고 정속으로 구동되는 엘리베이터를 구현하기 위해 DC Motor를 선택하여 정속제어 하였다. Atmega16을 통한 모터를 제어하기 위해서 PWM 방식을 사용하였으며, 또한 음성인식을 위해 음성인식 모듈을 사용하였고, LCD를 이용해 동작에 따른 상태를 표시하여 나타내었다. 음성인식 엘리베이터는 총 10층을 승강할 수 있도록 Atmega16을 통해서 구현하

였다.

3. 구동부

ATmega16을 통한 출력결과를 LCD를 통해서 나타내고, 승강기의 층수변화를 나타내었다.

3.1 LCD

LCD를 통해서 층수변화를 나타내고 모터가 정속제어를 수행하고 있음을 나타내었다. LCD 전용 컨트롤러는 HD44780을 사용하였다. LCD 모듈의 사양을 간략히 요약하면 <표2>와 같다.

<표2> HD44780 LCD 전용 컨트롤러

핀 번호	기 호	기 능
1	Vss	0V
2	VDD	5V
3	VL	VR 10k
4	RS	H:데이터, L:인스트럭션
5	R/W	H:리드, L:라이트
6	E1	H:인에이블 신호
7	D0	•데이터 버스
8	D1	
9	D2	•4비트사용 시:D4~D7만 사용 상위 4비트
10	D3	
11	D4	•하위4비트리드/라이트
12	D5	
13	D6	•8비트사용 시: D0~D7 사용
14	D7	
15	A	LCD 백 라이트 전원
16	K	

3.2 구동 모터

사용한 DC 모터의 제어원리와 ATmega16을 통한 PWM 방식을 적용하였다. DC 모터 제어란 모터의 속도 및 위치를 원하는 대로 조정하는 것을 의미한다. DC 모터 제어를 위해서는 모터 축 또는 기어 축의 위치(회전각도)를 측정할 수 있는 엔코더가 부착되어 있어야 한다. 제어기는 이 엔코더로 부터 들어오는 정보로부터 모터의 속도 및 회전각을 계산하고 이를 바탕으로 모터

에 가해지는 전압을 조절하게 된다. 모터 제어부는 엔코더의 펄스를 세어 12비트 또는 16비트의 펄스 수로 바꾸어 주는 계수기와 이 계수기 값으로 속도 및 위치를 계산하여 모터에 가할 전압을 결정하는 논리회로, 그리고 적정 전압과 전류를 공급하여 구동시키는 driver로 구성된다. 제어부의 구성은 크게 두 가지로서, 즉, 각각의 부품으로 하나의 제어부를 구성하는 방법과, 제어부 전체를 담당하는 하나의 칩으로 구성하는 방법이다. 후자의 경우는 시판되고 있는 모션컨트롤러라는 전용 칩을 사용하는 것이다. 이 칩은 단지 주 CPU에서 속도 및 위치를 입력하면 칩 자체에서 일정한 샘플링시간으로 스스로 모터를 제어하기 때문에 주 CPU의 부하를 크게 줄일 수 있다는 장점이 있다. 하지만 음성센서를 이용하여 일련의 동작을 수행하기 때문에 전자를 선택하였다.

ATmega16의 모터제어는 PWM 방식을 사용하였다. PWM 방식은, DC 모터제어에서 공급전압 변화방식이 아닌 구동스위칭펄스의 듀티 비(T1/T)를 변화시키는 방식으로 디지털방식의 제어가 가능하다. 듀티 비를 증가시키면 모터에 흐르는 전류의 시간평균값이 증가하며, 그 결과 속도가 증가한다. 그림4와 5에서 시스템의 플로차트와 프레젠테이션 동작 및 제어 과정을 나타냈다.

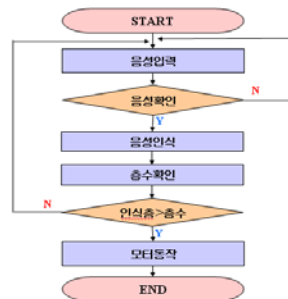


그림4. 플로우차트

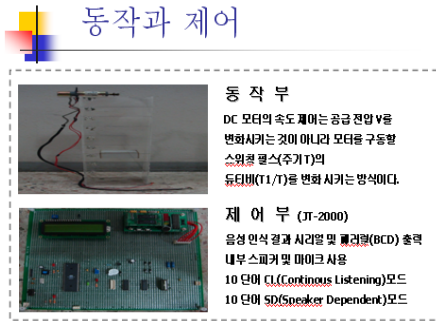


그림5. 프레젠테이션

III. 결과 및 고찰

음성인식 결과를 Atmega16에 전송하고, 그 결과에 따라서 모터를 제어하였다. 하지만 작동 중 개선되어야 할 점이 크게 2가지가 발견되었다. 첫째, 음성인식 모듈을 이용해 음성을 인식할 때, 모듈이 인식할 수 있는 음성은 처음 저장된 사람의 목소리가 한계이다. 즉, A 음성을 저장 시키면 A의 목소리만 인식하고, B의 음성을 저장시켜놓으면 B의 목소리만 인식할 수 있게 된다. 둘째, 모터의 부정확도이다. 엘리베이터의 동작은 모터를 이용해 모터의 회전수와 엘리베이터 외관의 1층간의 간격의 비례도를 계산해 구현되는데, 모터회전수에 따라 이동하는 거리가 정확하지 않다는 것이다. 이러한 현상이 발생하는 원인은 모터를 완벽히 제어하더라도 모터 내의 결함 또는 모터와 외관 사이의 접촉부분의 틈이 생긴 때문으로 판단된다.

기능에 대한 부가 설명을 하면, 엘리베이터는 사람이 쉽게 위층과 아래층으로 이동할 수 있도록 도와주는 기계장치이다. 사람의 안전과 관계되는 장치이므로 신뢰성을 갖추는 것이 중요함은 당연한 일이다. 따라서 우리는 부하에 따라 변하지 않고 정속으로 구동되는 엘리베이터

를 구현하기 위해 DC Motor를 정속제어 하였다. PWM 또한 음성인식을 위해 음성인식 모듈을 이용하였으며 LCD를 이용해 동작에 따른 상태 표시를 나타냈다. 음성인식 엘리베이터는 총 10층을 구현하였다. 사용한 음성인식 모듈이 인식할 수 있는 음성 단어의 개수가 열개이었고, 또한 자리 숫자 층에서 두 자리 숫자 층을 모두 구현할 수 있음을 보여주기 위해서 10층까지의 작동을 구현하는 시작으로 시연하였다. 그림6은 자연과 인간이 어울리듯 서있는 중국 호남성 장가계의 313m 높이의 백룡엘리베이터로써, 심산 중에서 규암기둥을 잡고 서있는 모습은 자연속 인간기술의 총아인양 우뚝 솟아있다.



그림6. 백룡엘리베이터

IV. 결론

음성은 문자 정보보다 약 4배정도 더 빠르고, 유선이나 무선 통신을 통한 원거리 정보전달 입력이 가능하기 때문에 다른 정보 수단보다도 훨씬 빠르고 수월하다.

따라서 컴퓨터가 인간의 음성명령을 수행하는 것과 같이, 기계 설비인 엘리베이터를 음성으로 명령하여 작동시킴으로써 육체적 기능으로 수행할 수 없는 특징적 장점을 구현하여 편의성 외에도 여러 가지 응급/긴급 상황의 대처기능을 실현시켜 확인하였다.

음성인식 감응 동작의 결과, 음성신호 특징에 따라 다소 차이가 있으나 지정인의 명령에 따라서 음성인식 감응에 의한 승강구간의 층 간 운행이 자연스럽게 작동됨을 확인하였고, 운행 중 긴급 상황발생 시 육체적 동작에 따른 긴급 명령 전달체계의 지연시간 발생을 음성신호 전달방식으로 대응함으로써 신속하게 지연시간을 단축/제거하고, 승강운행 중 돌발적으로 발생할 수 있는 응급/긴급 상황을 효과적으로 대처할 수 있음이 확인되었다.

특히 과중하게 운행되고 있는 관광지의 고층고속엘리베이터는 항상 긴급 상황이 발생할 수 있는 개연성은 매우 높다. 그림6에서 보듯, 높은 규압기둥에 붙여 세워진 백통엘리베이터는 자연속의 인간기술의 총아라고 해도 될까? 중국 장가계의 경우, 깎아지른 절벽바위산 중에서 산 중턱까지 무려 313m 높이의 고층 엘리베이터가 설치되어 고속 운행되고 있으며, 이러한 실례는 얼마든지 많이 있다. 특히, 시즌에는 과중한 운행의 정도가 매우 심각하여 안전성에 의문을 두게 한다. 긴박하게 짧은 시간 속에서 발생하는 기내의 긴급 상황에 실질적인 물리적 대응이 어려운 점을 감안하여, ‘특정 음성의 소리를 질러서라도’ 위험을 전달할 수 있는 긴급 대응체계 구축에도 일익을 담당해야한다. 음성인식방식으로 화자종속방식을 채택한 결과, 화자의 음성신호에 잡음이 가해질 경우는 인식오류가 발생할 수 있으므로 향후 보강이 요구된다.

따라서 이러한 오류에 대한 보완책이 강구되어, 향후 실험의 정성적 평가방식에서 정량적 측정과 평가가 동시 수행, 확보될 수 있는 체계를 구축하고, 현재 음성인식센서 장비의 고가문제도 풀고, 고층 동굴승강기의 적용과 함께 저가 활용도를 높이는 방안도 함께 강구되어야 하겠다.

참고문헌

- 1] C 언어를 이용한 AVR ATmega16 마이크로 컨트롤러/송봉길-성안당
- 2] C 언어를 이용한 AVR ATmega 128 마이크로 컨트롤러/송봉길-성안당
- 3] <http://cafe.daum.net/sayvoice> 음성정보기술 인식
- 4] <http://cafe.daum.net/avr8051>
- 5] AVR 마이크로프로세서 길잡이/신대섭-세화
- 6] <http://www.ktechno.co.kr>
- 7] <http://www.daehwagm.co.kr>
- 8] L. R. Rabiner, R. W. Schafer, DigitalProcessingofSpeechSignals,PrenticeHall,1978
- 9] Xuedong Huang, Alex Acero, Hsiao-wuen Hon, SpokenLanguageProcessing,Prentice Hall PTR, 2001
- 10] Herbert Gish and Michael Schmidt, “Text-Independent Speaker Identification”, IEEE Signal Processing Magazine, October 1994
- 11] Douglas A. Reynolds and Richard C. Rose, “Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models”, IEEE Trans. Speech and Audio Processing, Vol. 3, No.1, January 1995