

## 자동차 잡음환경에서의 음성인식시스템

김수훈\*, 안종영\*\*

### 요약

자동차 ECU(Electronic Control Unit)는 날이 갈수록 더욱 복잡해지고 많은 기능을 요구하고 있다. 대표적으로 power windows switch, LCM(Light Control Module), mirror control system, seat memory 등 운전자 편의 시스템이 개발되어 양산 중에 있다. 또한 현재 업계에서 많은 연구개발이 진행되고 있는 운전자 편의를 위한 DIS(Driver Information System)도 있다. 하지만 이러한 시스템을 운전 중 조작하게 되면 많은 위험이 따른다. 따라서 본 논문에서는 이러한 자동차 편의장치를 음성으로 조작 가능한 음성인식 시스템을 구현하였으며 자동차 잡음환경에서 인식을 향상 위한 전처리 필터를 적용하여 양호한 인식결과 얻었다.

## Speech Recognition System in Car Noise Environment

Soo-Hoon Kim\*, Jong-Young Ahn\*\*

### Abstract

The automotive ECU(Electronic Control Unit) becomes more complicated and is demanding many functions. For example, many automobile companies are developing driver convenience systems such as power window switch, LCM(Light Control Module), mirror control system, seat memory. In addition, many researches and developments for DIS(Driver Information System) are in progress. It is dangerous to operate such systems in driving. In this paper, we implement the speech recognition system which controls the car convenience system using speech, and apply the preprocessing filter to improve the speech recognition rate in car noise environment. As a result, we get the good speech recognition rate in car noise environment.

Keywords : Speech recognition, Filter, Driver convenience system, Car noise environment

### 1. 서론

현재 자동차업계에서는 자동차 개발환경에서 많은 변화를 시도하고 있다. 각 자동차 업체는 차별화된 경쟁력 강화 방안으로 디자인의 차별화와 첨단기능장착을 주요목표로 삼고 있다. 특히, 첨단기능장착은 자동차 전자장치의 증가를

의미한다. 즉 차량용 ECU(Electronic Control Unit)의 증가를 의미한다고 해도 과언이 아니다. 현재 자동차 부품에서 전자장치가 차지하는 비율은 약 25%정도 이지만 향후 2010년 이후 자동차에서 차지하는 비율이 40%이상 될 것으로 예상하고 있다[1][2]. 이에 따라 운전자가 조작해야 될 편의 장치 또한 증가될 것이다. 만약 운전자가 운전이외의 편의장치 조작에도 관여 한다면 분명 안전상의 문제점도 야기되리라고 사료된다. 따라서 음성을 통해 자동차 편의장치를 조작할 수 있다면 가장 이상적이라고 할 수 있다.

음성인식기술은 1960년대 이래로 많은 발전을 해 왔지만 아직도 기계에 의한 연속음성인식, 합성에는 많은 어려움과 연구해야 될 부분이 많이 남아 있다. 최근에는 고립단어 기반의 상용제품도 등장하고 있다. 지금의 상용제품이 나오기 까

※ 제일저자(First Author) : 김수훈  
접수일:2009년 02월 27일, 완료일:2009년 03월 20일  
\* 부천대학 모바일통신과  
shkim@bc.ac.kr  
\*\* DMC(주) 기술고문  
▣ 본 연구는 2007년도 부천대학 교비지원 연구비에 의하여 지원된 연구의 결과임

지는 상당한 시간이 걸렸으며 많은 기업들이 더 많은 음성인식 기술을 수용하려고 준비 중이고 다양한 접목이 시도되고 있다. 음성인식에는 특정 화자를 인식하는 화자중속방식과, 불특정 화자를 인식하는 화자독립방식으로 나눌 수 있으며 인식단어에 따른 고립단어인식과 연속단어인식으로 나누어진다[1]. 화자중속에는 화자인증, 핵심어인증으로 나눌 수 있다. 이와 같은 음성인식은 현재는 상용화 된 제품들도 있으며 자동차에도 적용하기위해 많은 시간과 노력을 하고 있는 실정이다. 일반 환경에서도 음성인식제품들이 어려움을 겪고 있는 부분은 바로 주변 잡음이 가장 큰 원인으로 분석되고 있다. 특히, 차량 이동시의 음성인식은 주변 환경 잡음에 의한 영향으로 음성 인식률이 감소한다. 따라서 음성인식 시장을 활성화하는데 있어 가장 큰 문제점은 잡음이라고 할 수 있다. 이러한 잡음을 제거하기 위하여 잡음제거를 위한 기술개발이 활발하게 이루어지고 있다. 현재 음성인식에서의 잡음처리 기술은 크게 음성향상(speech enhancement), 특징보상(feature compensation), 모델적응(model adaptation)과 같이 세 가지로 구분된다. 또한, 일반적으로 음성인식 알고리즘은 크게, HMM(Hidden Markov Model)[1][4], 신경망으로 나누어지나 현재는 HMM이 좋은 성능으로 자리매김 하였다[2][3][5]. 하지만 주변 잡음으로 인한 인식률 감소는 아직도 많은 연구가 이루어지고 있는 실정이다[8].

본 논문에서는 차량용 편의시스템 제어를 위한 음성인식 시스템을 구현하였고, 차량잡음으로 인한 인식률 감소를 개선하기 위하여 전처리 필터를 사용하였다. 2장에서는 음성인식시스템, 잡음처리필터, ECU 인터페이스 및 실험결과에 대하여 설명하고, 3장에서 결론을 기술하였다.

## 2. 구현 및 결과

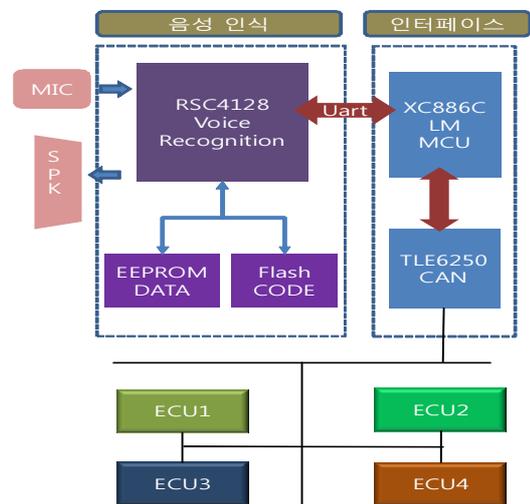
### 2.1 음성인식시스템

본 논문에서는 Sensory사의 음성인식 프로세서인 RSC-4128을 이용하여 음성인식 시스템을 구현하였다.

RSC-4128은 별다른 학습과정이 필요 없는 화자독립과 사용자의 학습과정이 필요한 화자중속 및 화자인증의 기능을 가지고 있다. 화자독립

의 경우 HMM과 신경망 기술이 접목된 하이브리드기반 기술을 적용하고 있고, 화자중속 및 화자인증의 경우는 DTW(Dynamic Time Warping) 패턴매칭 기술을 이용한다. 20비트 어드레스와 8비트 데이터 인터페이스로 외부메모리를 사용할 수 있다[2]. (그림 1)은 ECU 음성인식시스템의 개요이다. 음성인식 모듈에서 인식과정을 수행하고 인식결과를 Uart를 통해 인터페이스 모듈의 MCU(XC886CM)로 결과를 송신하고 그 데이터를 CAN BUS를 통하여 주변 ECU로 송신 한다. 각 ECU는 송신 받은 결과를 수행하여 결과적으로 음성인식을 통해 차량 편의시설을 제어 할 수 있게 된다[2][6].

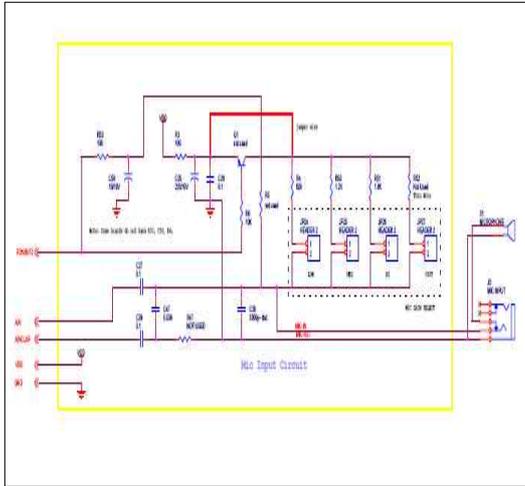
본 논문에서는 화자중속방식으로 22개의 명령에 대하여 2회 학습 후 인식 가능하도록 하였다.



(그림 1) ECU 음성인식시스템

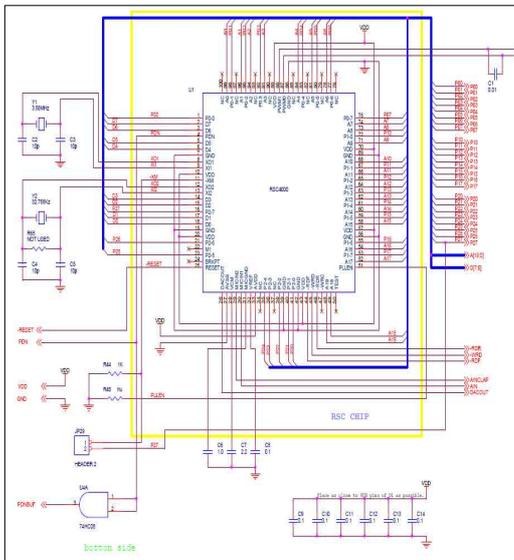
그리고 인터페이스는 HI CAN을 적용 하였으며 Infineon사의 MCU(XC886CLM)와 CAN transceiver(TLE6250)를 사용하였다. 그림2는 아날로그 음성 입력 부로써 4가지의 Pull-Up 저항으로 민감도를 조정할 수 있도록 설계 되어 있다. 저항 값이 클수록 민감도가 높아진다. 음성 입력 부는 음성 입력 시스템에서 음성을 받아들이는 부분으로 중요한 부분의 하나이다.

R, C값을 조정하여 민감도 및 입력특성을 조정할 수 있다.



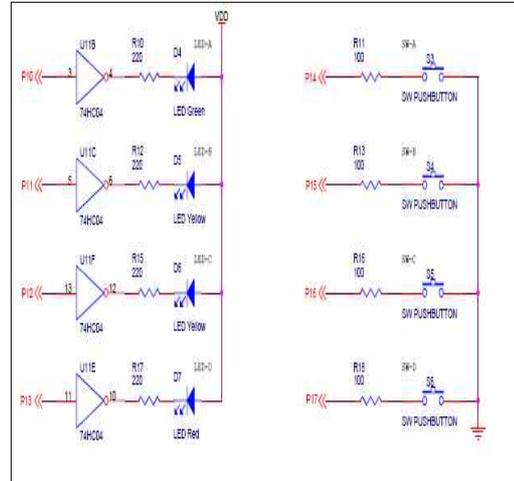
(그림 2) 음성입력 부

(그림 3)은 시스템의 Main Processor부이며 입력되는 음성신호에 대해서 음성인식과정을 수행하여 음성인식 여부를 진행하는 부분으로 결과 값을 I/O Port를 통하여 출력한다.



(그림 3) Main Processor

음성인식알고리즘(DTW)을 수행하는 부분으로 시스템에서는 가장 주요한 부분이라고 할 수 있다. 인식결과는 (그림 4)와 같이 LED(Red, Green, Blue)를 통해 알 수 있으며, 또한 음성 확인메시지를 스피커를 통해 확인 할 수 있다.



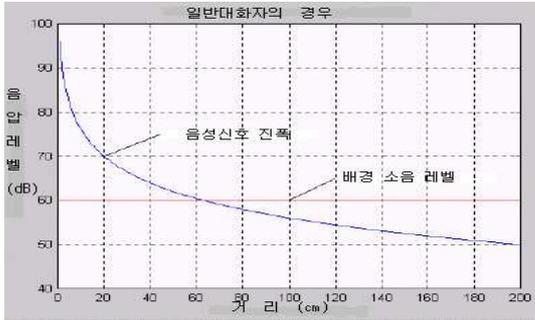
(그림 4) 시스템작동 및 확인 부

통상적인 음성인식시스템과 유사하게 4개의 Push Button을 통하여 음성인식, 화자학습, 데이터삭제기능을 가능하게 하고 있다[10].

본 논문에서 구현한 음성 시스템은 화자의 음성만을 인식하여 결과를 수행하는 시스템으로 화자의 2회 발성(학습)에 의해 플래시 메모리에 저장되며 버튼 제어에 의해 시스템이 가동된다. 프로그램에서는 매칭 레벨을 1~5까지 줄 수 있으며 이 값은 보이스 키의 인증레벨의 개념으로 사용할 수 있다. 키 레벨을 올리려면 매칭레벨을 올려서 값을 조정하여 사용할 수 있다. 레벨 값을 너무 올리면 보이스 키로 사용하기 힘들 정도의 인증과정을 거칠 수 있으므로 유의 하여야 한다.

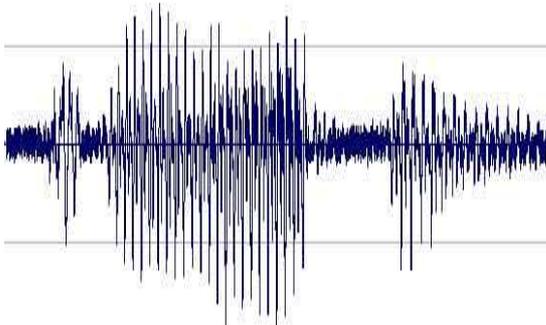
## 2.2 잡음처리

음성인식 시스템에 근거한 계측기는 일반적으로 20dB 이상의 신호 대 잡음비를 필요로 한다. 일반적인 음성인식을 위해서는 통상 마이크의 특성을 많이 고려하게 된다. 적용 아이템에 따라서 지향성, 무지향성을 고려하여 사용하게 되는데 중요한 것은 마이크의 지향성이 아니라 원음 대비 잡음 즉, 신호 대 잡음비(SNR)를 고려해야 한다. (그림 5)에서 알 수 있듯이 거리에 따라서 음압레벨이 급속히 감쇄함을 알 수 있다.



(그림 5) 잡음에 대한 음성신호의 음향 신호 대 잡음비

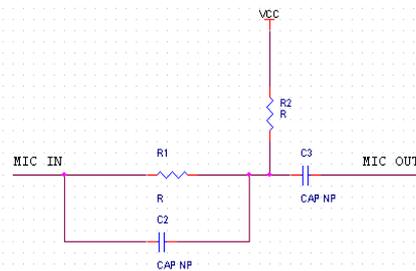
따라서 일반적인 음성인식의 경우 마이크와 발생되는 화자의 입과의 거리는 30cm 이내로 본다면 배경 잡음레벨과 음성과의 차이는 약 5~20dB 정도의 차이를 가진다고 할 수 있다. 상기와 같은 결과에서 볼 때 실제적으로 차이가 나는 잡음레벨을 어느 정도 없앨 수 있다면 주변 잡음을 제거할 수 있다고 할 수 있다. 기존의 입력음압레벨을 인위적으로 낮춘다면 음성의 레벨이 낮아지지만 배경잡음 역시 낮아지게 된다. 음성인식 시 가장 큰 문제가 되었던 배경잡음은 음원이 화자보다는 먼 거리에 위치해 있으므로 음성신호 진폭보다는 적다고 할 수 있다. 따라서 (그림 6)과 같이 잡음환경에서 화자가 발생한 음성은 마이크 입력 단에서는 분명 배경잡음보다는 발생음의 음압레벨이 크다고 할 수 있다.



(그림 6) 잡음환경에서의 음성신호

본 논문에서는 상기와 같이 음성과 배경잡음의 관계를 이용하여 입력음향을 전체적으로 진폭을 줄여서 음성감도를 둔화 시켜서 배경잡음의 레벨을 최소화하여 음성특징만을 추출 할 수 있도록 하드웨어적으로 필터를 설계하였다. (그림 7)과 같

이 마이크 입력 다음부터 콘덴서 및 저항의 병렬 필터를 구성하고 저항을 거친 DC성분 제거를 위하여 종단 필터용 콘덴서를 사용했다. 우선 초단 저항 및 콘덴서의 역할은 몇 가지가 있지만 입력음에 대해서 진폭을 낮추는 효과로 본 논문에 있어서 음향 스케일링 역할을 해주는 주요 포인트이다. 특히, 저항(R1)과 콘덴서(C2)의 용량 조절은 음향 스케일링을 조정하는 중요한 요소이다. 통상적으로 유사하게 사용되는 회로 이지만 본 논문에서의 특징은 기존에 사용되었던 저항(R1)의 값과 콘덴서(C2)의 용량 값을 조정하여 일반적인 음향레벨 값보다 더 이하로 조정하여 사용 하였다는 점이다. 이렇게 1차적으로 진폭조정이 되면 종단 콘덴서(C3)를 거치면서 DC성분의 잡음을 다시 제거한다. 결과적으로 진폭이 원음에 비해 약 4~5배 정도 줄어들게 되는데 음성신호는 물론 배경잡음도 같이 줄어들게 되므로 음성신호대비 배경잡음에 대한 변별력을 높일 수 있다.

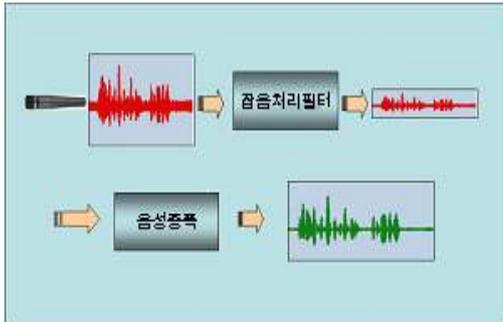


(그림 7) 잡음처리 필터

이렇게 1차적으로 진폭조정이 되면 종단 콘덴서(C3)를 거치면서 DC성분의 잡음을 다시 제거한다. 결과적으로 진폭이 원음에 비해 약 4~5배 정도 줄어들게 되는데 음성신호는 물론 배경잡음도 같이 줄어들게 되므로 음성신호대비 배경잡음에 대한 변별력을 높일 수 있다. 스케일링된 음성신호는 배경잡음레벨 이상인 음성신호만을 증폭시킴으로서 배경잡음을 거의 제거할 수 있다. 따라서 음성특징추출 및 음성인식 시 배경잡음을 (그림 8)과 같은 과정으로 제거한 다음 음성인식을 행하였다.

현재 차량잡음환경에서 잡음제거를 위하여 가장 많이 사용되고 있는 방법은 마이크입력부에 저역통과필터를 사용하는 방법이다. 하지만 이와 같은 방법으로 실험한 결과 평균 인식률은 30%

이하로 저조하였다. 하지만 본 논문에서 제안한 잡음처리방법을 사용한 후 현저히 향상된 인식 결과를 얻을 수 있었다.

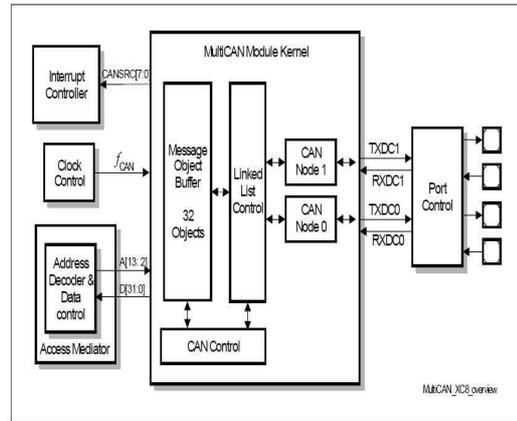


(그림 8) 잡음처리과정

### 2.3 ECU 인터페이스

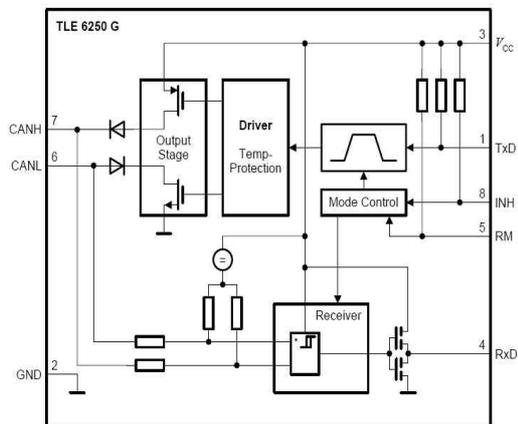
Robert Bosch사에서 개발한 CAN(Control Area Network)은 차량용 LAN의 표준 인터페이스 규격으로 현재 가장 많이 보급되고 있는 방식이다. 현재 각종 자동차 산업 분야로 적용범위가 확장된 방식으로 자동차뿐만 아니라 FA나 선박, 의료기기, 산업기기 등 다방면에서 도입이 진행되고 있다. 전송 속도는 최대 1M비트/sec이며 전송방식은 Event Trigger형으로 CAN에서는 1개의 노드가 통신경로를 접거하여 지연이 발생할 수 있다. CAN은 마이크로컨트롤러 간의 통신을 위해 설계되었으며 현재 자동차에서 증가하는 ECU간의 통신과 작업배선 저 잡음을 위해 설계되었다. 1987년 인텔에서는 82526이라는 표준으로 처음 CAN 라이선스를 소개하였고 1991년 Mercedes S-Class에서 처음 CAN을 적용하였다. 이때 발표된 CAN은 Specification 2.0으로 표준 11비트, 확장 29비트의 ID로 구성된다[9]. 본 논문에서는 Infineon사의 8비트 MCU XC886CLM과 CAN Transceiver인 TLE 6250을 사용하였다. XC886CLM은 ISO 11898 규격을 지원하고, MultiCAN 모듈 즉 2개의 Full-CAN 노드를 가지고 있으며 독립적인 32개의 메시지 오브젝트를 가지고 있다.

송신 2바이트 데이터 Length code를 사용하고 수신 3바이트 데이터 Length를 사용하였다. Baud Rate 500Kbps 그리고, Standard 11비트 ID를 사용하였다.



(그림 9) MultiCAN 모듈블럭다이어그램

TLE6250은 전송속도를 1M/bps 까지 높일 수 있는 High speed Can transceiver 이다. MCU 인터페이스를 위한 TX, RX핀을 가지며 CAN BUS Line CANH, CANL를 가지는 Dual line transceiver 이다.



(그림 10) TLE6250 Block Diagram

CAN 송신관련 함수는 다음과 같이 수행된다.

기본적인 타이머 인터럽트 함수를 이용하여 총 8바이트 전송이 가능하다. 물론 관련 함수는 main() 에서도 호출 가능하다. 먼저 데이터배열을 선언하고, 선언된 배열에 전송하고자 하는 데이터 값을 대입한 후 어드레스를 지정하여 데이터를 전송한다. 아래 코드는 2바이트 데이터송신 함수로 본 논문에 사용되어진 21개의 결과에 대한 전송을 충분히 고려하였다. 예를 들어 「창문 열어」의 전송에 대한 값은 0x11이다. 따라서

Transmit\_Data\_Array[0]=0x11; 값을 전송 하면 된다. 「단아」의 경우 Transmit\_Data\_Array[0]=0x12;를 전송하면 된다. 그리고 나머지 1바이트는 Temporary로 정의해 두었다. 통상 자동차에서는 8바이트 데이터를 사용하지만 본 논문에서는 2바이트를 음성인식결과에 대한 전송으로 사용 하였다.

```
void T01_viTmr0(void) interrupt T0INT
{
    unsigned char Transmit_Data_Array[2];
    Transmit_Data_Array[0] = 0x11;
    // 전송데이터
    Transmit_Data_Array[1] = 0x22;

    CAN_vWriteCANAddress(CAN_MODATAL0);
    CAN_DATA0 = Transmit_Data_Array[0];
    CAN_DATA1 = Transmit_Data_Array[1];
    CAN_vWriteEN(D0_VALID||D1_VALID);

    CAN_vTransmit(0); // Message object #0
}
```

CAN 수신 함수는 다음과 같이 수행된다. 수신 인터럽트발생 시 수신 오브젝트 어드레스를 지정하고 데이터를 수신한다.

```
void SHINT_viXINTR5Isr(void) interrupt XINTR5INT
{
    .....

    CAN_vWriteCANAddress(CAN_MODATAL16);
    CAN_vReadEN();
    Recieve_Data_Array[0] = CAN_DATA0;
    Recieve_Data_Array[1] = CAN_DATA1;
    Recieve_Data_Array[2] = CAN_DATA2;
    ....
}
```

**2.4. 실험결과**

차량용 ECU 제어를 위한 음성인식을 위하여 음성인식 전용 IC인 RSC-4128을 사용하여 화자 인식 실험을 하였다. 실험방법은 1명의 남성 화

자가 22개의 인식단어를 각각 2회 발성(학습)하여 플래시 메모리에 저장한 후, 10회 발성으로 인식률을 측정 하였다. 실험환경은 20Km/h 이하의 주행속도로 실험하였으며 일관성 있는 실험 결과를 위해 주변에 차량이 없는 장소에서 실험 하였다. 이러한 실험과정을 5명 화자에 대하여 각각 실시하였다.

<표 1> 인식률(%)

인식단어	화자				
	1	2	3	4	5
창문 열어	90	90	100	90	90
단아	100	100	100	90	100
창문 올려	90	80	90	90	90
정지	80	90	80	80	90
비상등	90	100	100	100	90
와이퍼	100	90	90	90	90
실내등	90	90	90	90	80
미등	90	100	90	90	90
좌측 깜빡이	80	70	70	80	70
우측 깜빡이	80	80	70	70	70
오디오	80	80	70	80	70
라디오	80	70	70	70	80
CD	80	90	90	70	80
DVD	80	90	80	80	80
볼륨 크게	80	70	70	70	80
볼륨 작게	70	70	80	70	70
전조등	90	90	90	90	90
상향등	90	80	80	90	90
에어컨	100	90	100	100	90
히터	100	80	90	100	90
네비게이션	100	100	100	90	100
핸즈프리	100	80	90	90	80

실험결과 화자 및 인식단어에 따라 조금씩 차이를 보이는데 그 이유는 완전 독립단어와 유사 단어의 차이와 각기 화자의 명령어 발생 시 생기는 음압레벨의 차이 때문이라고 사료되어진다. 「오디오」, 「라디오」, 「볼륨 크게」, 「볼륨 작게」와 같은 유사단어군 들이 「단아」, 「네비게이션」과 같은 비 유사단어에 비해 비교적 인식률이 떨어짐을 알 수 있다.

### 3. 결 론

본 논문에서는 음성인식 시스템을 구현하여 차량의 편의장치용 ECU제어를 위해 CAN 인터페이스와 접촉 시켜 보았다. 22개의 인식단어에 대하여 인식률을 산출하였고 차량용 인터페이스로 HI CAN BUS로 구현 해 보았다.

잡음차리필터의 경우 음성신호 대비 주변잡음이 약 10dB정도 이상의 음압레벨이 있어야 그 효과가 있다고 할 수 가 있으며 음성신호 대비 주변잡음의 음압레벨이 유사할 경우 주변 잡음필터의 효과가 떨어진다는 단점이 있다. 따라서 인식률 향상을 위해서는 화자가 다소 큰 소리로 명령어를 발성 하여야 한다. 실험결과에서 화자간의 인식률 차이도 상기와 같은 화자의 음압차이로 인한 결과로 추정된다. 이처럼 차량환경에서의 음성인식은 유사단어 및 주변잡음에 따라 많은 영향을 받음을 알 수 있다. 주변잡음의 경우 어느 정도 변별력을 향상시킬 수 있는 이론적인 접근이 가능하나 유사단어의 경우 변별력을 가지기에는 아직도 많은 노력이 필요한 현실이다. 향후 유사단어의 인식향상을 위하여 명령어 선정 시 중복 되는 단어선택을 배제하는 방법과 학습 시 변별력 있는 음절에 음압 차이를 두어 효과를 얻는 방법 등의 연구가 진행되어야 한다고 사료되어 진다.

### 참 고 문 헌

[1] WaldenC.Rhines, "System Approaches to Integration of Automotive Electronic Components", Automotive Electronics Conference, pp.7-26, 2006  
 [2] Seongsoo-Hong, "Technology Trends in Automotive RTOS and Component Middleware", Electric-Electronics Part Symposium, The Korean Society of Automotive Engineers, pp35-10, 2006  
 [3] Gordon E. Pelton, Voice Processing, pp90-102  
 [4] K-F.Lee and H-W.Hon, "Large-Vocabulary Speaker-Independent Continuous Speech Recognition Using HMM: The SPHINX System", proc ICASSP, pp3-126, 1988  
 [5] L.R.Bahl, P.F.Brown, P.V. de Souza, R.L.Meecher and M.A. Picheny, "Acoustic Markov Models used in the TANGORA speech recognition system", proc. ICASSP, pp497-500, 1988

[6] Sensory Speech 7 Technology, RSC-4X Evaluation Manual, 2003  
 [7] Raimund Ellinger, Burkhard Pollak, Thomas Pels, "Tasks and Process Steps to Develop and Evaluate Hybrid Electric Vehicles for Passenger Car to Heavy Duty Application", International Hybrid Electric Vehicle Workshop, 2007  
 [8] Jongsoon Jong, Younjeong Kyung, Seungho Choi, Hwangsoo Lee, "A Study on the performance improvement of speaker recognition using average pattern and weighted cepstrum", Korean signal processing conference, pp179-183, 1995  
 [9] Ki-Ho Kang, Bong Kyun Cho, "Design of a CAN-based Distributed Controller for BCM in Integrated Power/Signal MUX of Car", Electric-Electronics Part Symposium, The Korean Society of Automotive Engineers, pp.114-121, 2006  
 [10] 김수훈, 안종영, 로봇제어용 음성인식시스템 구현, 2004학년도 RT전략과제 개발사업 결과보고서, 2004.



**김 수 훈**

1990년 : 동아대학교 전자공학과 (공학사)  
 1992년 : 동아대학교 대학원 전자공학과(공학석사)  
 1999년 : 동아대학교 대학원 전자공학과(공학박사)  
 2001년~현 재: 부천대학 모바일통신과 부교수  
 관심분야 : DSP, 음성인식, 모바일콘텐츠



**안 종 영**

1993년 : 동아대학교 전자공학과 (공학사)  
 1996년 : 동아대학교 대학원 전자공학과(공학석사)

1996년~2000년: 현대전자 전장사업부  
 2001년~2003년: 아산기능대학 교수  
 2004년~2006년: 대성전기(주) 연구소 선임연구원  
 2006년~2009년: 창남(주) 기술연구소 수석연구원  
 2009년~현 재 : DMC(주) 기술고문  
 관심분야 : DSP, 음성인식, 디지털콘텐츠