

고속 주행중인 자동차 환경에서의 음성인식 연구

유봉근*, 이정기*, 김순협*, 박찬석**, 이순재**

* 광운대학교 컴퓨터공학과, ** 기아자동차 기술센터

A Study of Speech Recognition in a High Speed Automobile

Bong-Keun Yoo*, Jeong-Gi Lee*, Soon-Hyob Kim*, Chan-Seok Park**, Soon-Jae Lee**

* Kwangwoon University, ** KIA Motors

ybk@explore.kwangwoon.ac.kr

요약

본 논문은 고속 주행중인 자동차 환경에서 운전자의 안전 및 편의성을 위하여, 음성인식 기술을 이용한 각종 차량 편의장치를 제어하는 것으로, 운전자와 자동차와의 Man Machine Interface(MMI) 구조로 구성되었다. 이 시스템은 주행중인 자동차 환경에서 보조적인 스위치의 조작없이 상시 음성의 입, 출력이 가능하도록 하며, band pass filter를 이용하여 잡음 환경에서 자동으로 정확하게 음성구간 검출(End Point Detection)을 하게 하였다. Reference Pattern은 Dynamic Multi-Section(DMS)[1] 모델을 사용하였고 차량의 속도에 따라 자동으로 잡음환경에 강인한 모델을 선택하도록 하였으며, 음성의 특징 파라미터와 인식 알고리즘은 Perceptual Linear Predictive(PLP)[2] 13차와 One-Stage Dynamic Programming(OSDP)[3]을 사용하였다. Off-Line 실험 결과 고속 주행중인 자동차 환경(경부, 중부 고속도로 80km/h 이상)에서 자주 사용하는 차량제어 명령 33개에 대하여 화자독립 82.47%(중부 고속도로), 85.93%(경부 고속도로), 화자종속 93.43%(경부 고속도로)의 인식율을 구하였고, 30-70km/h 상태로 주행 중인 시내도로 On-Line 실험에서는 화자독립 92.89%, 화자종속 94.44%의 인식율을 구하였다. 또한 고속 주행중인 차량에서 카폰, 핸드폰 사용으로 인한 사고를 줄이기 위하여, 음성으로 전화를 걸 수 있도록

하는 Voice Dialing 기능도 구현하였다.

I. 서론

자동차의 편의 장치가 증가하면서 이들의 조작에 따른 운전자의 집중도 감소로 인해 교통사고의 위험성이 증가하고 있다. 최근에는 손 및 시선 집중의 부담이 없는 음성인식 기술을 적용하여 운전자의 안전 및 편의성을 확보하려는 시도로써 차량용 음성인식 장치 관련 연구 개발이 활발히 진행되고 있다. [4][5][6]

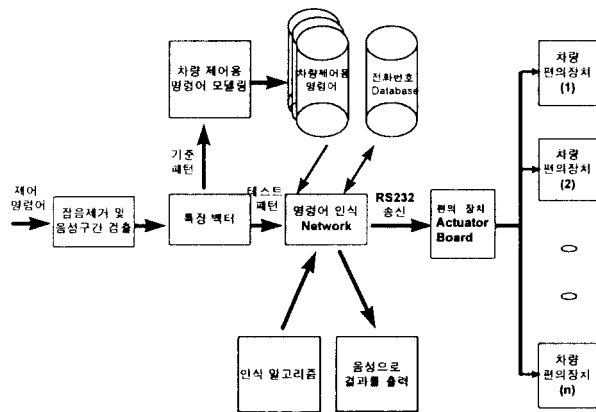
차량용 음성인식 장치는 운전자에게 자동차 문화의 부가적인 서비스를 제공할 뿐만 아니라, 자동차에 대한 친근감을 부여한다. 위와 같은 차량용 음성인식 장치는 주행중인 차량의 어떤 보조적인 스위치 조작 없이 동작이 가능하여야 하며, 손 보다 빠르게 차량의 편의장치를 제어할 수 있어야 한다. 그리고 어느 누구나 사용이 가능하여야 하며, 안전성을 위하여 높은 인식률이 요구된다. 인식률 향상을 위해서는 주행 중 가변적인 잡음환경에 강인한 음성인식 알고리즘의 개발이 필수적이다.

II. 본론

2.1. 시스템 개발 환경

고속 주행중인 자동차 환경에서의 음성인식 연구

음성입력은 편 타입 진방향성 콘덴서 마이크를 통해 이루어지며, 11.025kHz 샘플링 주파수로 이산화되어 16bits로 양자화 된다. 이 과정은 일반적인 PC용 사운드 카드를 통해서 이루어지며 노트북 PC상에서 비주얼 C++를 이용하여 음성신호처리 및 알고리즘이 실시간으로 구현된다. 음성취득 및 실험은 중부고속도로와 경부 고속도로에서 이루어지며 사용된 차량은 기아자동차의 포텐샤이다. 본 논문에서 사용한 단어는 비상등, 실내등, 에어컨등을 제어하기 위한 차량제어 명령어 22단어와 Voice Dialing을 위한 숫자음 11단어이며, 이들 단어는 주행중인 차량에서 화자와 마이크 거리를 30cm정도로 두고 취득하였다. 【그림 1】은 차량용 음성인식 장치의 구성도를 보인다. 먼저 주행중인 자동차에서 화자가 발생한 제어 명령어는 Band Pass Filtering 알고리즘을 이용하여 잡음제거 처리 후 제어명령어 즉 음성구간을 검출(End Point Detection)한다. 검출한 제어명령어는 PLP계수를 사용하여 특징벡터를 구하고 구해진 특징벡터 값을 이용하여 기준 패턴(Reference Pattern)과 테스트 패턴(Test Pattern) 처리를 한다.



【그림 1】 차량 음성인식 장치 구성도

이렇게 인식 알고리즘(OSDP)을 사용하여 구해진 인식 결과는 스피커를 통하여 음성으로 출력하고, RS232 케이블을 이용하여 편의장치 Actuator Board로 전송한다. 그리고 편의장치 Actuator Board는 전송된 결과를 이용하여 각종 차량 편의장치를 기동한다.

【그림 1】 시스템 구성도에서 표시된 바와 같이, 사용하는 기준 패턴은 차량 제어용 명령어 데이터베이스(DB)와 사용자가 음성으로 진화를 걸 수 있게 하는 Voice Dialing을 위한 전화번호부 데이터베이스 2가지가 있는데, 전자는 각종 차량 편의장치를 기동하기 위한 것이고, 후자는 사용자가 등록한 전화번호를 사용하여 음성으로 진화를 걸 때 사용하기 위한 것이다.

2.2. 음성구간 자동 검출

주행중인 자동차 환경에서의 잡음특성은 거시적인 특성일 뿐, 자동차 종류는 물론, 차 속도, 도로 요철 상태, 주변 환경에 따라 달라지며, 특히 음성신호의 처리단위인 프레임(Frame, 10~20ms)의 관점에서 보면 음성신호에 영향을 주는 가변적인 잡음요소들은 음성인식에 있어 큰 제한 요소가 된다.

결국 잡음이 존재하는 환경에서의 음성인식은 1) 전처리에 의해 잡음을 제거한 후 음성인식을 수행하는 방법, 2) 이미 존재하는 잡음에 대해 강한 알고리즘을 사용하는 방법, 3) 1), 2)의 장단점을 취한 혼합 방법이 있을 수 있다. 본 논문은 3)의 방법으로서 잡음제거를 통하여 음성구간(잡음제거된 음성구간)을 효과적으로 검출한다.

본 논문에서는 【그림 2】처럼 주행중인 차량에서 실시간으로 입력된 음성을 검출하기 위하여 Band Pass Filtering 처리하여 노이즈를 제거한 후에 1차로 음성구간과 노이즈구간을 포함한 음성을 구하고, 그리고 다시 2차로 음성구간만을 구한다.



【그림 2】 음성구간 검출 전체 블록도

음성구간과 북음구간을 분리해 내는데 가장 널리 이용되는 방법은 영교차율(Zero Crossing Rate)과 단구간 에너지(Energy)이다.[7] 하지만 주행중인 자동차의 배경잡음에서는 영교차율과 단구간 에너지만으로는 끝심검출을 한다는 것이 매우 어렵다. 본 논문에서는 영교차율과 단구간 에너지를 이용하기 전에 식 (1)을 이용하여 잡음을 제거한다. 그리고 음성검출 구간을 식(2), (3), (4), (5), (6), (7)을 이용하여 구한다.

$$y(n) = \sum_{k=1}^{10} x(n-k)h(k) \quad (1)$$

식 (1)에서 $x(n)$ 은 배경잡음이 섞인 음성 신호이고, $h(k)$ 는 필터 계수이다. 이때 $y(n)$ 은 필터링한 출력값을 가르킨다.

$$ZCR(i) = \sum_{n=1}^{N-1} \text{sgn}(y_{n+1}(i) \cdot y_n(i)) \quad (2)$$

$$E(i) = \sum_{n=1}^N |y((i-1) \cdot N + n)| \quad (3)$$

식 (2)와 (3)은 영 교차율과 단구간 에너지를 구하는 식으로 i 는 프레임의 번호이고 $ZCR(i)$ 는 i 번째 프레임의 영교차율 수이다. N 은 샘플수 128을 가르키며, $E(i)$ 는 i 번째 프레임의 에너지 값의 합을 의미한다. 본 논문에서는 음의 크기를 감소시킨 삼음과 음성구간을 좀 더 정확하게 구별하기 위하여 식 (2), (3)에 제곱을 하여 식 (4), (5)를 구한다.

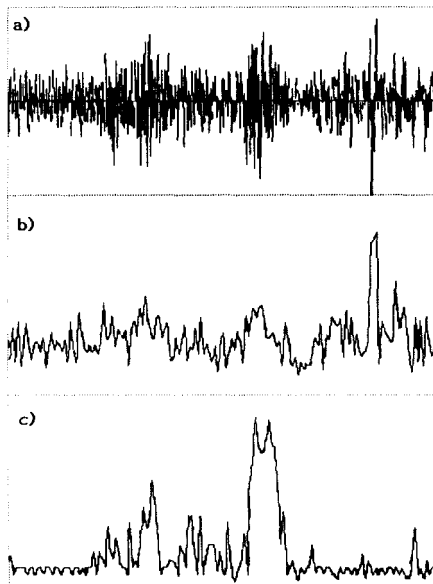
$$\overline{ZCR} = ZCR(i)^2 \quad (4)$$

$$\overline{E} = E(i)^2 \quad (5)$$

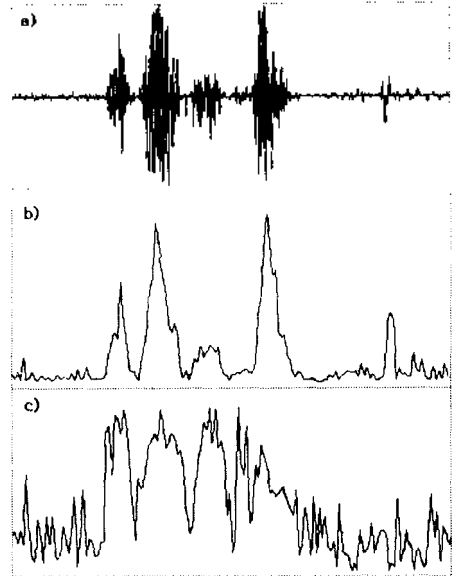
$$\overline{E} \times a < E(i)^2 \quad (6)$$

$$\overline{ZCR} < ZCR(i)^2 \quad (7)$$

기준 Threshold는 주행중인 자동차 잠음에 적용시키기 위하여 일정 시간마다 재조정 하고, 식 (6), (7)의 \overline{E} , \overline{ZCR} 은 식 (4), (5)에서 구한 Threshold의 에너지 값과 영교차율이며 E , ZCR 은 실시간으로 입력되는 잠음 또는 잠음섞인 음성구간의 에너지 값과 영교차율을 가르킨다. 음성구간의 시작부분을 구하기 위해서는 식(6), (7)을 이용한다. 식(6), (7)을 동시에 연속적으로 4프레임이상 만족하면 음성구간으로 간주한다. 만약 이 조건을 만족하지 않으면 잠음구간으로 처리한다. 여기에서 a 는 실험치이다.



a) 100km/h 전후의 속도에서 취득한 '비상등 커'
 b) a)신호에 대한 단구간 에너지
 c) a)신호에 대한 ZCR
 【그림 4】 중부 고속도로에서 취득한 음성 데이터



a) 【그림 4】의 잠음제거된 파형
 b) a)신호에 대한 단구간 에너지
 c) a)신호에 대한 ZCR
 【그림 5】 잠음제거된 음성신호에 대한 영교차율과 단구간 에너지

음성의 끝부분 확인은, 음성이 끝난 이후에 일정한 시간 동안에 잠음이 계속되면 즉 식 (6), (7)을 만족하지 않으면 음성입력이 끝난 것으로 한다.

【그림 4】는 중부 고속도로에서 취득한 데이터로 단구간 에너지와 ZCR을 보이고 있으며, 【그림 5】는 이 데이터에 대하여 잠음을 제거한 파형과 단구간 에너지, ZCR 이다.

2.3. 음성모델 구성

80km/h 이내에서 주행중인 차량의 음성 모델은 Idle 상태로 구성된 데이터가 Idle 상태와 주행중인 상태에서 취득한 데이터의 혼합보다 인식율이 높게 나타났고, 고속 주행중인 차량에서는 후자의 모델이 인식율이 높게 나타났다.[8]

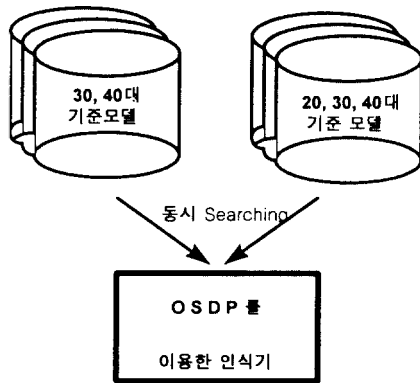
따라서 본 논문은 Idle 상태와 40km/h 상태로 주행중인 차량에서 취득한 데이터를 혼합하여 【표 1】과 같이 모델을 생성하고 실험을 한다.

【표 1】 모델 구성을 위한 환경

모델		차속도									
		1	2	3	4	5	6	7	8	9	10
40km/h	발음	2	1	2	3	1	3	3	3	0	0
	횟수	2	3	3	3	2	2	1	0	3	2

고속 주행중인 자동차 환경에서의 음성인식 연구

그리고 인식 실험은 【그림 6】 과 같이 2개의 모델을 사용한다.[8]



【그림 6】 화자적응을 위한 기준 모델에 대한 구조

2.4. 인식결과

【표 2】 는 【표 1】 에서 선정된 모델에 대하여, 1프레임의 길이를 23.2ms로 하고 특징벡터는 PLP 13차, DMS 20차로 실험한 결과이다.

【표 2】 고속 주행중인 차량환경에서의 인식율
(단위 : %, OFF-Line 실험)

고속도로	모델	1	2	3	4	5
중부(화자독립)		77.49	80.95	80.95	80.30	80.09
경부(화자독립)		84.42	85.93	85.28	84.42	84.20
경부(화자종속)		90.91	93.43	92.42	92.93	90.91

고속도로	모델	6	7	8	9	10
중부(화자독립)		78.79	77.49	76.41	82.47	80.09
경부(화자독립)		84.20	83.55	81.39	85.06	85.93
경부(화자종속)		90.91	87.88	85.86	92.42	92.42

위의 실험은 Off-Line 실험이며 인식실험에 참여한 단어수는 462단어(중부 고속도로, 화자독립, 남자 3인), 462단어(경부 고속도로, 화자독립, 남자 3인), 198단어(경부 고속도로, 화자종속, 남자 1인)이다.

중부 고속도로가 이처럼 인식율이 나쁜 원인은 도로 상황으로 볼 수 있다. 왜냐하면 경부 고속도로는 아스팔트 도로인데 비해, 중부 고속도로는 시멘트 도로이므로, 주행중인 차량의 바퀴와 노면사이에 의하여 상당한 잡음이 발생해 잡음에 의한 에너지 값이 높기 때문이다.

【표 3】 은 30 ~ 70km/h 상태로 주행중인 서울 시내도로 On-Line 실험으로 화자독립 92.89%의 인식율을 보이고 있다.

【표 3】 주행중인 시내도로에서 화자독립 인식율
단위 : % (인식율(인식 참여 단어 수))

화자	인식율 (인식에 참여한 단어 수)
A (30代)	96.97 (99)
B (30代)	96.36 (165)
C (30代)	91.67 (132)
D (30代)	93.18 (132)
E (30代)	94.95 (99)
F (40代)	84.85 (165)
전체 인식율	92.89 (759)

2.5. Voice Dialing 기능

자동차 안에서의 카폰, 핸드폰 사용이 증가하면서 이들의 조작에 따른 운전자의 집중도 감소로 인해 교통사고의 위험성이 높다. 이는 전화번호를 누를 때 인지도의 약화와 시야 장애, 그리고 자동차의 제시간 지연등이 사고의 원인이라고 할 수 있다. 본 논문에서는 이러한 문제점을 해결하기 위하여 음성으로 전화를 걸 수 있도록 하는 Voice Dialing 시스템을 연구하였다.

특히 이 기능은 Man Machine Interface 구조로 구성되었으며, 전화번호를 등록하는 전화번호 등록기능과, 전화번호를 삭제하는 삭제기능, 그리고 등록되어 있는 모든 전화번호를 사용자에게 음성을 통하여 출력함으로써 사용자가 전화번호등록 및 삭제, 전화통화등을 편리하게 할 수 있도록 도와주는 등록된 전화번호 목록의 출력기능으로 나누어진다.

2.6. 편의장치 Actuator Board 구현

편의장치의 제어는 비상등, 에어컨(히터), 실내등, Power Window 편의장치로 제한하며 그 외의 명령어는 '인식어 Display 패널'의 LED ON/OFF를 통하여 확인한다. Actuator Board는 PC상에서 인식된 음성 명령어를 정해진 '1'에서 '33'까지의 숫자 형태로 전송 받아서 편의장치를 구동하거나 인식 명령어 Display 패널의 LED를 ON/OFF시킨다.

III. 결론

본 논문에서는 Band Pass Filter를 통해 잡음을 제거하고, 영교차율과 단구간 에너지를 이용하여 자동 음성구간 검출을 구현하였다. 영교차율과 단구간 에너지의 기준 값을 일정한 시간마다 잡음레벨에 의해 자동적으로 보정된다. VQ는 DMS 모델을 사용하였고, 잡음환상에 강인성을 갖도록 하기 위해 기본적인 배경잡음과 주행중인 차량의 배경잡음을 적절하게 사용하여 기준모델을 구성하였다. 화자의 적응력을

갖도록 하기 위해 30, 40代로 구성된 기준모델과, 20, 30, 40代로 구성된 기준모델 2가지로 모델을 구성하였고, 인식실험시에는 2개의 기준모델을 동시에 Searching 한 후 인식결과가 높은 명령어를 선택하도록 하였다. OSDP 인식 알고리즘으로 Off-Line 인식 실험을 행한 결과 고속 주행중인 자동차 환경(경부, 중부 고속도로 80km/h 이상)에서 자주 사용하는 차량제어 명령 33개에 대하여 화자독립 82.47%(중부 고속도로), 85.93%(경부 고속도로), 화자종속 93.43%(경부 고속도로)의 인식율을 구하였다.

PC에서 인식된 음성 명령어는 68HC11을 이용한 Actuator Board와의 RS232 통신을 통하여 비상등, 실내등, 에어컨(Blower), Power Window를 구동하고 인식된 명령어는 Display 패널에 표시되도록 구현하였다.

[참고문헌]

[1] 변용규, "DMS 모델을 이용한 단독어 인식에 관한 연구", 박사학위 논문, 광운대학교, 1990. 12

[2] H. Hermansky, "Perceptual Linear Predictive(PLP) Analysis of Speech" J. Acoust. Soc. Am. 87(4), pp1738 ~ 1752, April 1990

[3] Hermann Ney, "The Use of a One-Stage Dynamic Programming Algorithm for Connected Word Recognition", IEEE Transaction on Acoustics, Speech, and Signal Processing, Vol. ASSP-32, No.2, pp263 ~ 271 April, 1984

[4] 이기철, "차량소음에 강한 고립단어 음성인식에 관한 연구", MS Thesis, KAIST, 1995

[5] A.NOLL, "Problem of Speech Recognition in Mobile Environments", ICSLP90, Vol.2, pp1133 ~ 1136, 1990

[6] Chafic MOKBEL, Ge'rard CHOLLET, "An Improved Noise Compensation Algorithm for Word Recognition in the Car", ICASSP91, Vol.2, pp925 ~ 928, May, 14-17

[7] L. R. Rabiner, M. R. Sambur, "An Algorithm for Determining the End Points of Isolated Utterances", The Bell System Technical Journal, Vol.54, No.2, pp297 ~ 315, February 1975

[8] 유봉근 외 "주행중인 자동차 환경에서의 음성인식 연구", 한국음향학회 학술발표대회 논문집 제17권 제1(s)호, pp47 ~ 50, July 1998