1999년도 한국음향학회 학술발표대회 논문집 제18권 제1(s)호

차량내에서의 음성인식에 관한 연구

박정훈*, 임형규**, 김종교*

* 전북대학교 전자공학과
** 서남대학교 전산정보학과

A Study on Speech Recognition inside the Car

Jeong-Hoon Park*, Hyung-Kyu Im**, Chong-Kyo Kim*

* Dept. of Electronics , Chonbuk National University.
** Dept. of Computer and Information., Seonam University

요 약

본 논문은, 자동차에서 발생할 수 있는 다양한 형태의 잡음이 섞인 음성을 대상으로, 잡음에 강인한 파라미터들을 사용하여 인식기들을 구축하였으며, 이들 파라미터를 비교 평가하였다.

실험에 사용된 음성 데이터는 차종, 속도, 도로 환경, 라디오 ON/OFF, 창문 개폐여부 등 다양한 잡음 환경에서 수집하였다.

실험에서 비교된 파라미터는 MFCC(Mel-Frequency Cepstral Coefficient)와 PLP(Perceptually Linear Prediction)이며, 각각의 파라미터에 대해서 MKM(Modified k-mean)을 이용하여 코드북을 작성하였고, DHMM(Discrete Hidden Markov Model)을 인식알고리즘으로 사용하였다.

실험 결과로서, 아스팔트 도로에서 창문을 닫고, 라디오를 켜지 않은 상태에서 60km/h로 주행시 96.25%로 가장 높은 인식률을 얻었고, 고속도로에서 창문을 열고 100km/h로 주행시에는 60%로 가장 낮은 인식률을 얻었다.

1. 서 론

자동차내부의 보조장치들의 동작을 음성을 사용하여 동작시키면 주행중 운전자의 불필요한 행동을 감소시킬 수 있게되며 기기의 사용의 편리성과 운행중의 안정성을 증대시키게된다. 그러나, 자동차에서의 음성인식은 자동차 내·외부의 잡음 환경으로 인하여 많은 어려움이 존재한다. 따라서, 자동차내에서 음성인식은 잡음에 강인한 특징 파라미터 선정과 인식 알고리즘의 구현, 그리고 잡음 처리 기술이 중요하다.

각 특징 파라미터에 대한 인식률를 비교하였으며, 속도와 도로환경의 변화 그리고 라디오를 켠 경우와 켜지 않은 경우의 인식률을 평가하였다. 그리고 단어별 인식 실험을 하였다.

2. 잡음 데이터 베이스 구축

실험에 사용된 데이터는 가장 대중적인 차종(현대 소나타EF, 삼성 SM520)을 이용하여 녹음하였으며, 데이터 수집환경은 표1에서와 같이 차종, 속도, 도로환경, 라디오 ON/OFF, 창문의 개폐여부 등에 따라 수집하였다.

표1. 잡음 데이터베이스구축을 위한 환경

	수집환경	세부항목		수집환경	세부항목
1	자동차 2사	삼성	4	Radio ON/OFF	ON
		현대			OFF
2	주행속도	60km/h	5	도로상태	고속도로 (아스팔트)
		80km/h			국도 (아스팔트, 시멘트)
		100km/h			
3	주행중 창문의 개폐	열림	6	발성위치	운전석
		닫힘			조수석

주행중인 차량과 잡음이 없는 실험실 환경에서 DB를 구축하였다. 차량내에서 음성인식을 위한 차량 명령어 20개와 고립 숫자음 14개를 선별하였다. 표2와 같이 차량운행제어에 대한 명령어는 완벽한 인식을 보장할 수

없으므로 차량운행에 직접적인 관련이 없는 보조장치에 대한 단어를 중심으로 선정하였다.

표2. 차량 명령어에 대한 단어의 선정

	명령어		고립숫자음
1	비상등 켜	1	영
2	비상등 꺼	2	일
3	실내등 켜	3	이
4	실내등 꺼	4	삼
5	라디오 켜	5	사
6	라디오 꺼	6	오
7	소리 높여	7	육
8	소리 낮춰	8	칠
9	창문 열어	9	팔
10	창문 닫어	10	구
11	열선 켜	11	십
12	열선 꺼	12	백
13	에어콘 켜	13	천
14	에어콘 꺼	14	만
15	히터 켜		
16	히터 꺼		
17	왼쪽 깜박이		
18	오른쪽 깜박이		
19	와이퍼 동작		
20	와이퍼 멈춤		

데이터는 자동차 내부의 운전석과 조수석의 차광판 중앙에 위치한 Omni-directional fixed-charge 마이크를 통하여 Portable-DAT에 녹음한다. 녹음된 데이터는 Workstation 환경에서 ESPS(Entropic Signal Processing System)를 통하여 9600Hz로 샘플링 되고 16bits로 양자화된다.

수집된 데이터를 스펙트로그램으로 분석한 결과 여러 가지 환경에서 대체로 그림 1과 같이 잡음이 300Hz~500Hz에서 분포하므로 High Pass Filter(HPF)를 통과시켜서 잡음환경에서 발성음에 대해서 300Hz이하의 잡음은 제한한다.

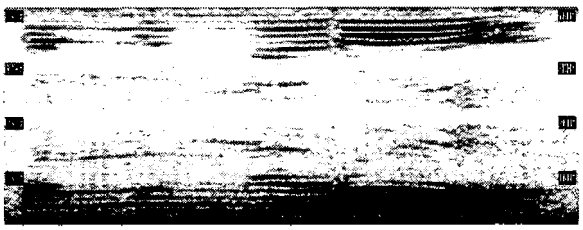

그림 1. 아스팔트도로, 창문닫고, 60km/h,
'창문열어'의 스펙트로그램

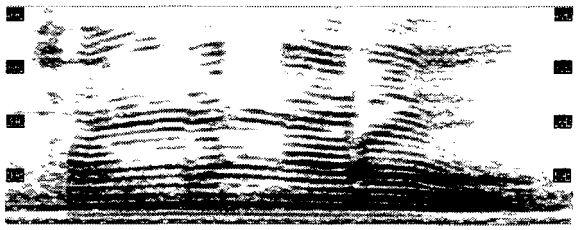

그림 2. 아스팔트도로, 창문닫고, 60km/h,
'창문열어'의 스펙트로그램

그림 1과 그림 2는 HPF의 통과전과 통과후의 스펙트로그램이다. 300Hz이하의 주파수 대역이 제한되었음을 볼 수 있다.

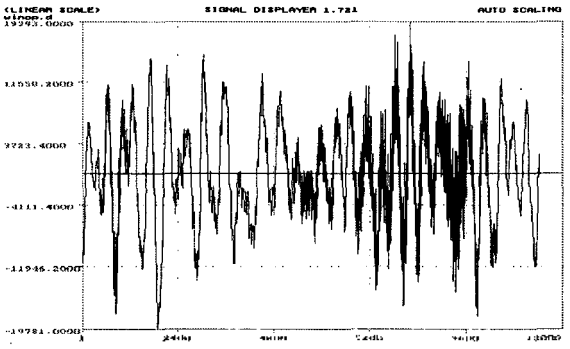

그림 3. 아스팔트도로, 창문닫고, 60km/h,
'창문열어'의 파형

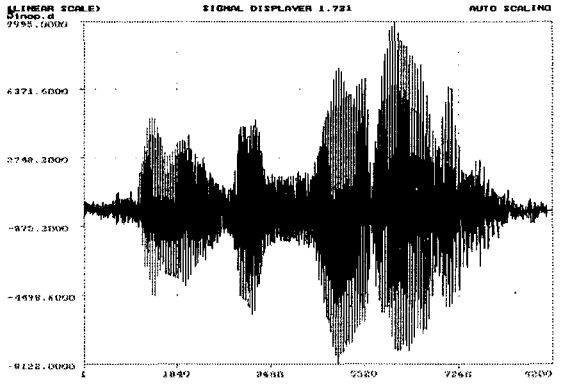

그림 4. 아스팔트도로, 창문닫고, 60km/h,
'창문 열어'의 파형

그림 3과 그림 4는 HPF를 통과하기 전의 파형과 통과 후의 파형을 나타낸 것이다. 통과 전은 파형이 육안으로 드러나지 않으나 통과 후에는 음성구간이 나타나게 된다.

3. 파라미터 추출

다양한 잡음이 존재하는 차량내에서 음성인식은 잡음에 강인한 파라미터의 추출이 중요하다.

3.1 PLP 분석

PLP는 발성된 음성을 귀의 인지능력에 맞춘 분석방법으로 일반적인 성도 모델링을 위한 LPC에 사람이 음성을 인지하는 방법론을 첨가한 것이다. 귀는 각 주파수에 따라 들려오는 강도가 틀리며, 들을 수 있는 주파수의 범위도 한정되어 있다. PLP가 기존의 LPC와 다른 점은 사람의 청각적 특성을 고려한 Critical-band 적분과 Equal-loudness 전처리과정 그리고, Cubic-root 압축방식이다.

3.2 MFCC

이전 프레임에서 추출된 특징과 이후의 프레임에서 추출된 특징간의 차를 새로운 특징 파라미터로 이용함으로써 전·후의 음소에 의해서 발생된 음소의 변화를 특징 파라미터에 포함시킬 수 있다.

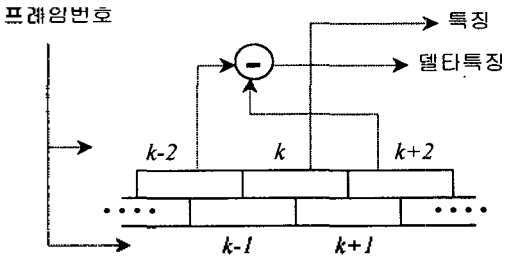

그림 5. 특징 파라미터 추출 방법

그림 5는 본 논문에서 사용하는 특징추출방법을 나타내었다. 실시간 인식 시스템 구성의 경우를 고려함으로써, 실험에서 사용되는 모든 특징은 독립된 프로세스가 처리하도록 하는 것을 가정하여 하나의 함수로 구현되었으며, 20ms폭의 프레임에 대한 특징으로 11차의 mel 켑스트럼과 1차의 에너지와 이의 델타mel-켑스트럼 12차로 구성된 총 24차의 특징을 이용하고 있다. 그림 6은 전체 특징 추출의 블록도이다.

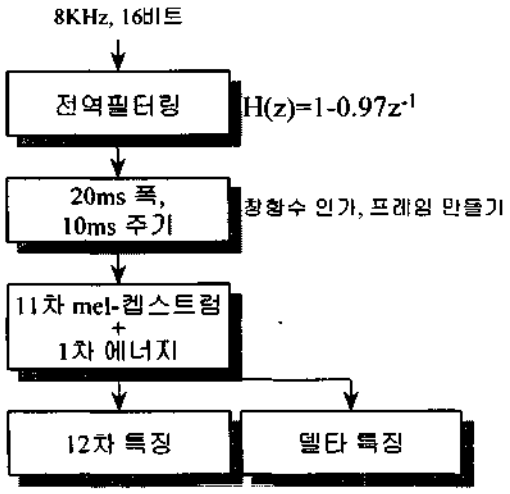

그림 6. 전체특징추출의 블록도

4. 인식 모델의 구성

차량내에서 20대 남성화자 2명이 차량 명령어 20단어와 고립 숫자음 14단어를 차종별, 주행 환경별로 구분하여 총 2244개의 어휘와 실험실 환경에서는 20대 남성화자 12명, 30대 남성 4명과 20대 여성화자 4명 총 21명이 714단어를 발성하였다. MFC파라미터의 현재값과 이전값과의 차를 합하여 VQ를 행하였다. 코드북 설계를 위한 VQ알고리즘은 MKM(Modified k-mean)을 이용하여 상태수 128을 지닌 하나의 코드북, 글로벌 코드북을 설계하였다.

Test 패턴의 구성은 Training 패턴에 참여하지 않은 데이터로 소나타EF와 SM520이 60km/h, 80km/h의 속도로 아스팔트 도로와 시멘트 도로를 주행한 경우, 고속도로 환경에서는 소나타EF의 80km/h로 주행한 환경과 100km/h의 속도로 두 차량이 모두 주행한 경우를 선택하여 총 10가지의 모델을 갖추었으며, 다시 10가지 모델에 대하여 라디오 ON/OFF, 창문의 개폐여부를 고려하여 모두 30가지의 모델을 구성하였다.

인식기는 단순 좌·우 모델을 이용한 DHMM 알고리즘을 사용하였다. 본 연구에서 인식하고자 하는 단어는 음소수가 대략 8~12개 정도로 구성된 차량 명령어와 1~3개로 구성된 숫자음이다. 그러므로 차량명령어는 상태수 10을 고립 숫자음은 상태 수 3으로 다르게 인식기를 구성하였다.

5. 실험결과 및 분석

잡음에 강인한 파라미터의 선정을 위하여 PLP와 MFC의 인식률을 비교, 분석하였다. 표3과 그림 7은 파라미터에 대한 환경별 인식률을 나타낸다.

표3. MFC와 PLP의 인식률 비교 단위(%)

파라미터	도로환경	CLW, ROF	CLW, RON	OPW, RON	Total
MFC	AR	96.25	92.5	86.25	**91.7**
	CR	91.7	88.6	74.3	**85**
	HR	93.3	90	60	**81.1**
PLP	AR	87.2	83.5	70	**80.3**
	CR	78	71.2	65	**71.4**
	HR	82.6	76.2	60.3	**73.3**

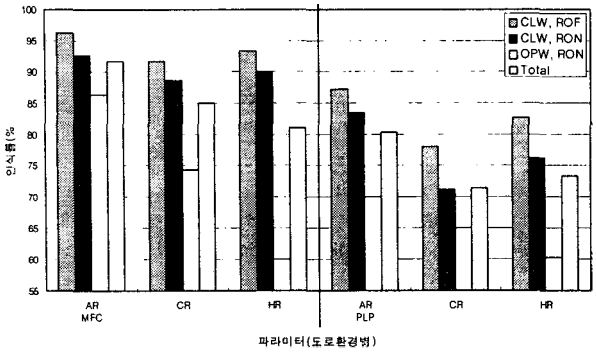

그림 7. MFCC와 PLP파라미터의 인식률 비교

여기서

AR : 아스팔트 도로(Asphalt Road)
CR : 시멘트 도로(Cement Road)
HR : 고속도로(High-way Road)
CLW : 창문 폐(Closed Window)
OPW : 창문 개(Opened Window)
RON : 라디오 ON(Radio On)
ROF : 라디오 OFF(Radio Off)

MFCC 파라미터는 평균 약 86%의 인식률을 보이나 PLP의 경우에는 74.9%로 MFCC가 PLP보다 11.1%정도의 높은 인식률을 나타낸다. 특히 창문을 닫고 라디오를 켜지 않은 경우의 인식률은 각 96.25%와 87.2%로 큰 차이가 있음을 알 수 있다. 그러나 창문을 열고 100km/h로 주행한 경우에는 인식률의 차이가 나타나지 않으며 대단히 저조하다.

표 4는 차량 명령어에 대한 인식률을 나타낸다.

표 4 차량 명령어와 차내 환경별 비교 (단위 : %)

	단어목록	CLW, ROF	CLW, RON	OPW, ROF	Total
1	비상등 켜	100	100	100	100
2	비상등 꺼	90	90	60	77
3	실내등 켜	90	90	80	87
4	실내등 꺼	100	100	70	87
5	라디오 켜	90	100	90	94
6	라디오 꺼	100	80	70	80
7	소리 높혀	100	90	70	87
8	소리 낮춰	90	90	80	87
9	창문 열어	90	90	70	84
10	창문 닫어	80	80	70	77
11	열선 켜	90	90	70	84
12	열선 꺼	90	90	70	84
13	에어콘 켜	80	80	90	84
14	에어콘 꺼	100	90	80	90
15	히터 켜	90	80	60	77
16	히터 꺼	100	90	80	90
17	오른쪽 깜박이	90	100	80	90
18	왼쪽 깜박이	80	90	80	80
19	와이퍼 동작	100	100	90	97
20	와이퍼 멈춤	90	100	90	94
	Total	**92**	**90.5**	**74.5**	**86.2**

명령어 인식에서 '켜'와 '꺼'를 오인식하는 경우가 많이 발생하게 되었다. 단어의 길이가 길수록 인식률이 높음을 알 수 있다. 특히 단어 중에서 '와이퍼 동작'과 '와이퍼 멈춤'은 '동작'과 '멈춤'과 같이 확연히 구분되는 단어로 인하여 다른 단어들에 비하여 인식률이 높다. 또한 'ㅁ', 'ㅇ'과 같은 비음을 포함하고 있는 어휘는 비교적 높은 인식률을 보였으며, 'ㅅ', 'ㅊ'과 같은 파찰음을 포함한 어휘에 대해서는 반대의 결과를 보였다.

그림 8은 고립 숫자음에 대한 인식률을 나타낸다.

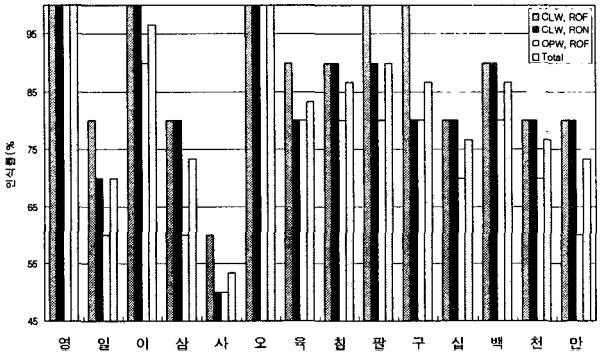

그림 8. 고립 숫자에 대한 인식률

고립 숫자음의 인식률이 차량 명령어의 인식률보다 약 4%정도 낮음을 볼 수 있다. 대체로 3음소로 이루어진 숫자음이 명령어에 비하여 정보량이 적기 때문이다. 특히 단어별로 볼 때 모음으로 구성된 '영', '이', '오'의 인식률이 대단히 높은 반면에 마찰음과 파열음으로 구성된 어휘는 인식률이 낮았다.

각 숫자음에 대해서는 '일'은 '이'과 '칠'로 많이 오인식을 하였으며, '삼'은 '사'과 '만'으로 오인식하였다.

다양한 속도를 주행하는 차량의 도로환경에 대한 인식률 비교를 그림 9에서 나타내었다.

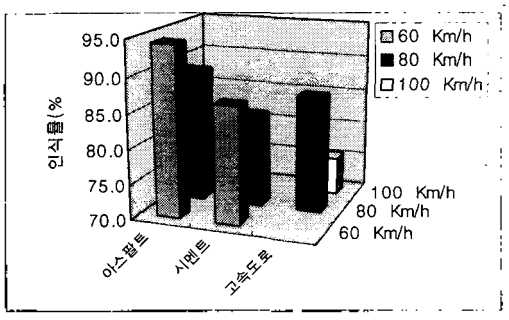

그림 9. 속도와 도로환경에 따른 인식률

속도별 인식률에서는 아스팔트 도로에서 60km/h로 주행한 경우가 평균 90.5%로 가장 높았으며, 100km/h로 주행한 고속도로에서는 평균 75.5%를 나타내었다. 시멘트 도로가 아스팔트 도로에 비해 주행중인 차량의 바퀴와 노면사이에 상당한 잡음의 발생으로 인하여 인식률이 저조하다.

6. 결 론

본 논문은 차량내에서 음성인식을 목적으로 하였다. 차량명령어 20개 어휘와 고립 숫자음 14개의 어휘로 구성되었다. 우리 나라에서 대중적인 차량을 이용하여 차종, 도로환경, 속도, 라디오 ON/OFF, 창문의 개폐여부에 따라 다양한 환경별로 실험하였다.

전처리 과정을 거친 음성데이터에 대하여 잡음에 강인한 음성 파라미터를 구하고 MKM을 이용하여 코드북 사이즈가 128인 글로벌 코드북을 작성하였다.

파라미터에 대해서는 MFCC(평균 86%)가 PLP(평균 74.6%)보다 차량내 음성인식에서는 우수한 인식률을 보였으며 정보량이 많은 차량 명령어가 고립 숫자음보다는 인식률이 높았다. 도로환경에서는 아스팔트 도로가 91.7%로 다른 도로환경에 비하여 높은 인식률을 보였으며, 속도는 60km/h로 주행시 90.5%로 가장 높은 인식률을 보였다.

차후 연구 방향은 인식기 구현 시 차량 명령어와 고립 숫자음을 인식하는데 상태 수가 다르므로 하나의 인식기로 구현하는 알고리즘의 연구와 인식에 소요되는 시간을 줄이기 위해 인식기 내에서의 연산량의 감소에 대한 노력과 실시간 구현을 위한 DSP(Digital Signal Processing) board를 이용하는 방법도 검토되어야 한다.

참 고 문 헌

[1] 김종교외, "전화망에서 실시간 음성인식", 신호처리 합동학술대회 논문집, 제8권 제1호, pp.775-778, 1995. 9. 23.

[2] 김종교외, "A Real-time Feature Extraction and Endpoint Detection of Speech in the Telephone Network," *Proceedings of th ICEIC'95*, pp. II-143-146, peoples Republic of China, 1995. 8. 7-12.

[3] R. W. Schafer and J. D Markel, Speech Analysis. *IEEE Press*, 1979.

[4] J. Picone, "Continuous Speech Recognition Using Hidden Markov Models," *IEEE ASSP magazine*, pp. 26-41, July 1990.

[5] Y. Linde, A. Buzo, and R. M. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Trans. on Comm.*, vol. COM-28, no. 1, jan. 1980.

[6] L. R. Rabiner, J. G. Wilpon, and F. K. Soong, "High Performance Connected Digit Recognition Using Hidden Markov Models," *IEEE Trans. on ASSP*, vol. 37, no. 8, Aug. 1989.

[7] L. R. Rabiner, J. G. Wilpon. " A Segmental K-means Training Procedure for Connected Word Recognition Based on Whole Word Reference Patterns," *AT&T Tech. journal*, vol. 65, no. 3, pp.22-31, May/June 1986.

[8] C. Myers and L. R. Rabiner, " A level Building Dynamic Time Warping Algorithm for connected Word Recognition," *IEEE Trans. on ASSP*, vol. ASSP-29, no. 2, pp.283-297, Apr.1981

[9] L. R. Rabiner, A. E. Rosenberg, and S. E. Levinson, "Considerations in Dynamic Time Warping Algorithm for Discrete Word Recognition," *IEEE Trans. ASSP* vol ASSP-26, no. 6, pp. 575-582, Dec. 1978.

[10] J. G. Wilpon, C. H. Lee, and L, R, Rabiner, " Connected Digit Recognition based on Improved Acoustic Resolution," *Computer Speech and Language*, no. 7, pp. 15-26, 1993.

[11] Thomas W. Dennison, Frank J. Malkin and Christoper C. Smyth, "The Effect of Helicopter Vibration on the Accuracy of a Voice Recognition System," Aerospace Technology Conference and Exposition Long Beach, California. Oct. 5-8, 1987.

[12] M. Shimotani, M. Hibino, T. Yamamoto, and T. Nonami, "A voice Recognizer for Car Telephone System," International Congress and Exposition Detroit, Michigan. Feb. 26-Mar.2, 1990.

[13] Kapul D. Gill. and Ashraf Jawaid, "Speech Recognition An Application for Quality Assurance in the Automotive Industry," Passenger Car Meeting and Exposition Dearborn, Michigan. Sep. 17-20, 1990.