1998년도 한국음향학회 학술발표대회 논문집 제 17 권 제 1(s) 호

주행중인 자동차 환경에서의 음성인식 연구

유봉근*, 이정기*, 김학진*, 이성권*, 김순협*, 박찬석**, 이순재**

* 광운대학교 컴퓨터공학과, ** 기아자동차 기술센터

A Study on Speech Recognition in a running automobile

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

* Kwangwoon University, ** KIA Motors

요약

본 논문은 자동차의 편의성 및 안전성의 동시 확보를 위하여, 보조적 스위치의 조작없이 상시 음성의 입,출력이 가능하도록 하며, band pass filter를 이용하여 잡음환경에서 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) 13차와 One Stage Dynamic Programming(OSDP)를 사용하였다. 주행중인 자동차 환경(30~70km/h)에서 자주 사용되는 차량제어 명령 33개에 대하여 화자독립 92.89%, 화자종속 94.44% 인식율을 구하였다. 또한 주행중인 차량에서 카폰, 핸드폰 사용으로 인한 사고를 줄이기 위하여 음성으로 전화를 걸 수 있도록하는 Voice Dialing 기능도 구현하였다.

Ⅰ. 서론

자동차의 편의 장치가 증가하면서 이들의 조작에 따른 운전자의 집중도 감소로 인해 교통 사고의 위험성도 증가한다. 최근에는 손 및 시선 집중의 부담이 없는 음성인식 기술을 적용하여 운전자의 편의성 및 안전성을 확보하려는 시도로써 차량용 음성인식 장치 관련 연구 개발이 활발히 진행되고 있다.[2][3][4]

차량용 음성인식 장치는 보조적인 스위치 도움없이 동작이 가능하여야 하며 안전성을 위하여 높은 인식률이 요구된다. 인식률 향상을 위해서는 주행중 가변적인 잡음환경에 강인한 음성인식 알고리즘의 개발이 필수적이다.

Ⅱ. 본론

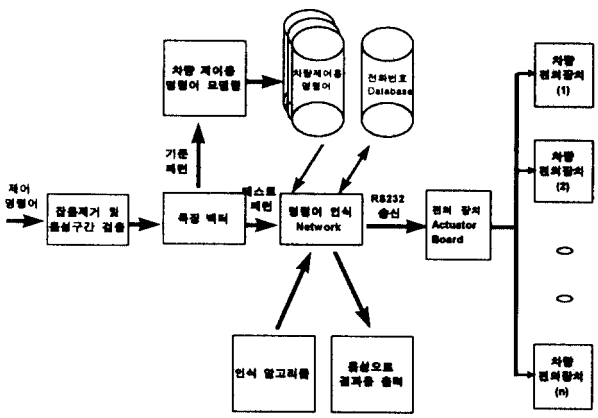
2.1. 시스템 개발 환경

음성입력은 핀-타입 전방향성 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 일반시내, 시외도로와 고속도로에서 이루어지며 사용된 차량은 기아자동차의 포텐샤이다. 【표 1】은 본 논문에서 사용한 차량제어 명령어와 Voice Dialing을 위한 숫자음 33단어를 나타내고 있으며, 이들 단어는 주행중인 차량에서 화자와 마이크 거리를 30cm정도로 두고 취득하였다. 【그림 1】은 차량용 음성인식 장치의 구성도를 보인다. 먼저 주행중인 자동차에서 사용자가 발성한 제어명령어는 Band Pass Filtering 알고리즘을 이용하여 잡음제거 처리 후 제어명령어 즉 음성구간을 검출(End Point Detection)한다. 검출한 제어명령어는 PLP 계수[5]를 사용하여 특징벡터를 구하고 구해진 특징벡터 값을 이용하여 기준 패턴(Reference Pattern)과 테스트 패턴(Test Pattern) 처리를 한다.

이렇게 인식 알고리즘(OSDP)[6]을 사용하여 구해진 명령어 결과는 스피커를 통하여 음성으로 출력하고, RS232 케이블을 이용하여 편의장치 Actuator Board로 전송한다. 그리고 편의장치 Actuator Board는 전송된 결과를 이용하여 차량 편의장치를 기동한다.

【표 1】 음성 명령어

구 분	명령어			
차량 제어 명령	1. 비상등 켜	2. 비상등 꺼		
	3. 실내등 켜	4. 실내등 꺼		
	5. 오디오 켜	6. 오디오 꺼		
	7. 소리 크게	8. 소리 작게		
	9. 다음 채널	10. 이전 채널		
	11. 라디오 스캔	12. 정지		
	13. 에어컨 켜	14. 에어컨 꺼		
	15. 히터 켜	16. 히터 꺼		
	17. 온도 올려	18. 온도 내려		
	19. 창문 올려	20. 창문 내려		
	21. 통화 시작	22. 통화 종료		
숫자음	23. 일	24. 이	25. 삼	26. 사
	27. 오	28. 육	29. 칠	30. 팔
	31. 구	32. 영	33. 공	



【그림 1】 차량 음성인식 장치 구성도

2.2. 음성구간 자동 검출

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

【그림 2】 음성구간 검출 전체 블록도

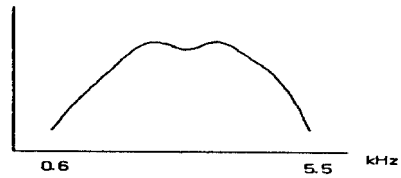
본 논문에서는 【그림 2】처럼 주행중인 차량에서 실시간으로 입력된 음성을 검출하기 위하여 Band Pass Filtering 처리하여 노이즈를 제거한 후에 1차로 음성구간과 노이즈구간을 포함한 음성을 구하고, 그리고 다시 2차로 음성구간만을 구한다.

2.2.1. Band Pass Filter

사운드 카드를 이용하여 입력된 데이터를 1400Hz에서 4200Hz 정도로 대역을 제한하며, 【그림 3】은 Band Pass Filter의 스펙트럼 형태로, 이와 같은 필터 계수와 배경잡음이 섞인 음성 신호의 필터링은 다음과 같다.

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

식 (1)에서 $x(n)$은 배경잡음이 섞인 음성 신호이고, $h(k)$는 필터 계수이다. 이때 $y(n)$은 필터링한 출력값이고, N 은 16차를 가르킨다.



【그림 3】 Bandpass Filter 스펙트럼 형태

2.2.2. 1, 2차 음성구간 검출

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

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

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

식 (2)에서 i는 프레임의 번호이고 ZCR(i)는 i번째 프레임의 영교차율 수이며, N은 샘플수 128을 가르킨다. 식 (3)에서 E(i)는 i 번째 프레임 에너지 값의 합을 의미한다. 본 논문에서는 음의 크기를 감소시킨 잡음과 음성구간을 좀 더 정확하게 구별하기 위하여 식 (2), (3)에 제곱을 한다.

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

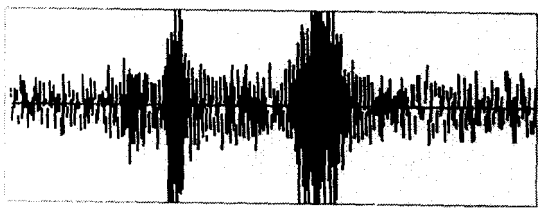
$$\overline{E} = E(i)^2 \quad (5)$$

$$\overline{E} < E(i)^2 \times 3.5 \quad (6)$$

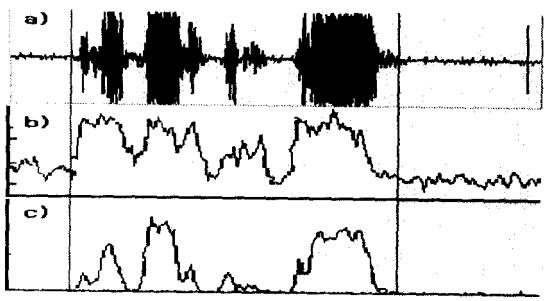
$$\overline{ZCR} < ZCR(i)^2 \quad (7)$$

가준 Threshold는 잡음에 적용시키기 위하여 1.5초마다 재조정 하고, 식 (6), (7)의 $\overline{E}$, $\overline{ZCR}$은 식 (4), (5)에서 구한 Threshold의 에너지 값과 영교차율이며 E, ZCR은 실시간으로 입력되는 잡음 또는 잡음섞인 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 잡음 구간으로 처리한다.

음성의 끝부분 확인은, 음성이 끝난 이후에 0.5~0.7초 동안에 잡음이 계속되면 즉 식 (6), (7)을 만족하지 않으면 음성입력이 끝난 것으로 처리한다.



【그림 4】 주행중인 환경에서의 '비상등 켜'



a) 【그림 4】의 band pass filter에의한 잡음제거
b) a)신호에 대한 ZCR
c) a)신호에 대한 단구간 에너지

【그림 5】 잡음제거된 음성신호에 대한 영교차율과 단구간 에너지

2.3. 음성모델 구성

【표 2】 모델 구성을 위한 환경

차속도/Order \ 모델		1	2	3	4	5	6	7
Idle	발음	2	2	2	2	3	2	3
40km/h	횟수	-	2	1	2	-	-	-
차수		13	13	13	13	13	13	13
Section		15	15	15	20	15	20	20

2.4. 인식결과

【표 3】은 【표 2】에서 선정한 모델에 대하여 실험한 결과로서, 1프레임의 길이는 23.2ms로 하였다. 표에서 볼 수 있듯이 PLP가 다른 특징 파라미터보다 인식율이 높은 것으로 나타났으며 모델 7번 즉 Idle상태에서 20대 남자화자 4인이 3번 발음한 것이 인식율이 가장 높은 것으로 나타났다. (인식에 참여한 단어 수 : 1287개)

【표 3】 20대 화자 4명의 40~80km/h에서의 종속 인식율 (단위 : %, OFF-Line 실험)

파라미터 \ 모델	1	2	3	4	5	6	7
LPC	83.03	85.47	86.92	85.41	86.02	88.42	89.04
Mel-Cep.	90.05	87.72	91.53	88.66	91.53	89.98	93.01
PLP	93.94	92.39	95.10	93.32	93.94	94.79	95.57

그러나, 고속도로(경부, 중부)에서 주행중인 (100km/h 전후) 차량 데이터를 저장한 후 OFF-Line 실험을 한 결과 Idle 상태보다는 Idle 상태의 데이터와 40km/h 주행시에서 취득한 데이터를 혼합한 모델이 인식율이 높게 나타났다. 【표 4】에서 인식실험에 참여한 데이터 수는 각각 165 단어와 198단어이며, 경부 고속도로에서는 화자종속, 중부 고속도로에서는 화자독립 실험으로 이루어졌다. 그리고 특징 파라미터는 잡음환경에서 인식율이 높은 PLP를 사용하였다.

【표 4】 고속도로 100km/h 전후에서의 인식율

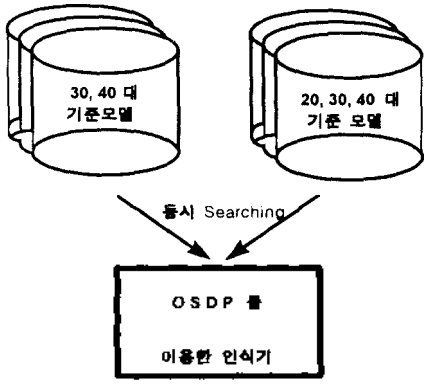
차속도 \ 모델	1	2	3	4	5	6	7
100 km/h(경부)	-	90.9	92.1	90.9	-	-	89.7
100 km/h(중부)	-	73.7	71.2	76.8	-	-	71.7

비록 화자종속(경부 고속도로) 및 화자독립(중부 고속도로) 실험이었지만 중부 고속도로가 이처럼 인식율이 나쁜 원인은 도로 상황으로 볼 수 있다. 왜냐하면 경부 고속도로는 아스팔트 도로인데 비해, 중부 고속도로는 시멘트 도로이므로, 주행중인 차량의 바퀴와 노면 사이에 의하여 상당한 잡음이 발생해 잡음에 의한 에너지 값이 높기 때문이다.

2.5. 화자독립 인식율 향상을 의한 처리

2.4.에서 구한 모델(PLP, 【표 2】의 모델7과 4)을 사용하여 주행중인 자동차(40-80km/h)의 실시간 환경에서 화자종속 및 화자독립 실험을 하였다. 화자종속 4인이 발음한 1320단어에서 94.45%의 인식율이 나왔지만, 20대-50대로 이루어진 10인(화자독립)이 발음한 924단어에서는 80.83% 결과가 나왔다.

화자독립 인식율 향상을 위하여 20대로 구성되었던 기준모델을 【그림 6】과 같이 20,30,40대로 구성된 모델과 30,40대로 구성된 모델 2개를 인식과정에서 동시에 searching 하도록 모델 구조를 바꾸었다.



【그림 6】 화자적응을 위한 기준 모델에 대한 구조

그 결과 실시간 환경에서, 30~70km/h 차량속도에서 30, 40代 화자 6인이 화자독립 92.89%(759단어)의 인식율을 보였고, 6인이 발음한 화자종속은 94.44%(990단어)의 인식율을 보였다.

2.6. 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 장해, 그리고 자동차의 제어시간 지연등이 사고의 원인이라고 할 수 있다. 본 논문에서는 이러한 문제점을 해결하기 위하여 음성으로 전화를 걸 수 있도록 하는 Voice Dialing 시스템을 연구하였다.

Voice Dialing 기능은 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.

2.7. 편의장치 Actuator Board 구현

편의장치의 제어는 '비상등 켜/꺼', '에어컨(히터) 켜/꺼', '창문 올려/내려', '실내등 켜/꺼'로 제한하며 그 외의 명령어는 '인식어 Display 패널'의 LED ON/OFF를 통하여 확인한다. Actuator Board는 PC상에서 인식된 음성 명령어를 【표 1】에서 정해진 '1'에서 '33'까지의 숫자 형태로 전송 받아서 편의장치를 구동하거나 인식 명령이 Display 패널의 LED를 ON/OFF시킨다.

Ⅲ. 결론

본 논문에서는 Band Pass Filter를 통해 잡음을 제거하고, 영교차율과 단구간 에너지를 이용하여 자동 음성 구간 검출을 구현하였다. 영교차율과 단구간 에너지의 기준 값은 1.5초마다 잡음레벨에 의해서 자동으로 보상된다. VQ는 DMS 모델을 사용하였고, 삼음환경에 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. 위와 같은 환경을 구축한 후, 실제 일반도로를 주행하면서 OSDP 인식 알고리즘으로 인식 실험을 행한 결과 30~70km/h 범위에서 화자독립 92.89%, 화자종속 94.44%의 인식율을 보였다. 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] 이기철, "차량소음에 강한 고립단어 음성인식에 관한 연구", MS Thesis, KAIST, 1995

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

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

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

[6] 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

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