자동차 환경에서 피치검출을 이용한 음성인식 연구

A study on speech recognition using pitch detection in a car-noisy environment

이정기, 유봉근, 김학진, 김순협

Jeong-gi Lee, Bong-keun Yoo, Hak-jin Kim, Soon-hyob Kim

광운대학교 컴퓨터공학과

Computer Engineering Kwangwoon University Korea

요 약

본 논문은 자동차의 편의성 및 안전성의 동시 확보를 위하여, 보조적 스위치의 조작없이 상시 음성의 입·출력이 가능하도록 하였고, 남성과 여성을 구별하기 위하여 피치검출법을 사용하여 속도별로 구분하였다. 또한, band pass filter를 이용하여 자동으로 잡음하에서 정확하게 음성구간 검출(End Point Detection)을 하게 하였다.

Reference Pattern은 DMS(Dynaminc Multi-Section)[1]모델을 사용하였고, 음성의 특징 파라미터와 인식 알고리즘은 PLP 13차와 One Stage Dynamic Programming(OSDP)를 사용하였다. 시내주행중인 자동차 환경에서 자주 사용되는 차량제어 명령어 30단어를 가지고 실험한 결과 40-80km에서 화자독립 남성 96%, 여성 94.4% 화자종속일 때 남성 97%, 여성 95%의 인식률을 얻을수 있었고 남성과 여성을 구분하므로 써 인식률을 향상 시켰다.

I. 서 론

인간과 기계간의 가장 쉬운 의사 소통 도구중의 하나인 음성이 자동차에 적용된다면 인간에게 더욱 편리할 것이다. 최근 음성인식에 대한 많은 관심을 갖고 여러 분야에 접목시키고 있는 실정이며국내·외에서는 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. 또한 여성 운전자들 증가하는 추세이기 때문에 남성실험에서 탈피하여 여성실험에도 큰 비중을 두어야 한다.

따라서 본 논문에서는 피치검출을 통하여 남성과 여성을 구분하여 실험을 하였다.

II. 본 론

II-1. 음성구간 자동 검출

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

【그림 1】 음성구간 검출 전체 블록도

본 논문에선는 【그림 1】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 따로 구한다.

II-2. Band Pass Filter

【그림 2】는 밴드 패스 필터의 스펙트럼 형태로써, 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 필터계수이다.

이때 $y(n)$은 필터링한 출력값 이고, N은 16차를 가르킨다.

【그림 2】 Bandpass Filter 스펙트럼 형태

Ⅱ-3. 음성구간 검출

음성구간과 묵음구간을 분리해 내는데 가장 널리 이용된 방법은 영교차율(Zero Crossing Rate)과 단구간 에너지(Energy)이다.[5] 하지만 주행중인 자동차의 배경잡음에서는 영교차율과 단구간 에너지만으로는 끝점검출을 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.

$$ZCR(i) = \sum_{n=0}^{N-1} sgn(y_{n+(i-1)*N}\, y_{n+1+(i-1)*N}) \quad (2)$$

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

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

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

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

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

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

기준 threshold는 잡음에 적응시키기 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.

Ⅱ-4. 피치검출

본 논문에서는 자동차환경에서 남성, 여성을 구별하기 위하여 피치검출을 사용하였다.

일반적으로 가장 많이 사용하는 피치검출법은 자기상관함수법인데 본 논문에서는 실시간으로 음성을 처리하다보니 속도가 빠른 AMDF법을 사용하였다.

자기상관함수(Autocorrealction)에서 $x(n)$과 $x(n+k)$곱 대신에 다음과 같이 절대값으로 정의된다.

$$AMDF(K) = \sum_{M=0}^{N-1} |x(m) - x(m-k)| \quad (8)$$

AMDF법은 자기상관 함수법에서 수행하는 곱 연산을 절대값과 차분으로 대신하기 때문에 상대적으로 빠르다는 장점을 가지고 있다. 이러한 이유로 실시간에 많이 적용한다.

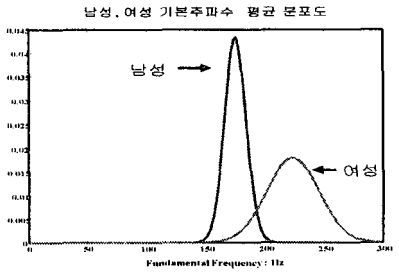
자기상관함수 $\Phi(k)$에서는 피치주기 배수에 최대 값을 이루지만, AMDF법에서는 피치주기 배수에 최소 값을 갖는다. 첫 번째 최소 값이 음성 프레임의 기본주파수 즉, 피치가 된다.

본 논문에서는 AMDF를 이용하여 Pitch를 추출해낸 다음 기본주파수(Fundamental frequency, F0)로 변환시킨다. 피치와 기본주파수는 같은 뜻으로 쓰이는데 피치는 시간 축에서 음성 파형의 주기를 찾아내는 것을 말하고, 이러한 피치를 주파수 축에서 해석한 것을 기본주파수라고 한다.

【그림 3】 은 1500cc 소형 승합차에서 차량제어 명령어를 발성했을 때 피치값의 평균과 표준편차 로 나타낸 정규분포도이다.

【그림 3】 은 자동차환경 에서 차량 제어명령어 를 남성과 여성으로 각각 속도별로 구분하여 실험한 결과 남성은 150-200Hz, 여성은 150-300Hz 범위에 분포하고 있었다.

본 논문은 자동차의 특수한 환경에 잡음이 섞인 피치주기 역시 함께 존재하고 있다.



【그림 3】 자동차 환경에서 남성, 여성 기본주파수 평균 분포도

II-5. 음성모델 구성

본 논문에서는 서울시내도로 및 내부 순환도로에서 님,여 각각 5명이 10번 발음한 데이터를 가지고 실험하였다. 이때 속도별로 실험한 결과 idel-40km, 40-80km, 80-100km로 속도별로 실험을 하면서 가장 적절한 모델을 다음과 같이 구성하였다.

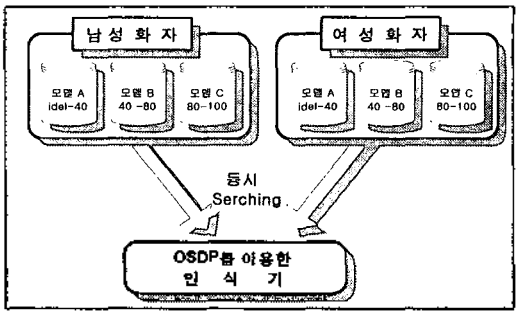
① Idle-40km, ② 40-80km, ③ 80-100km 모델

그리고, 차량 제어명령어는 【표 1】 과 같다.

【표 1】 차량 제어명령어 (30단어)

구 분	명 령 어
제 어 명령어	비상등, 와이퍼, 실내등, 오디오, 에어컨, 히터, 정지
	다음채널, 이전채널, 라디오스캔, 소리크게, 소리작게, 온도올려, 온도내려, 창문올려, 창문내려, 통화시작, 통화종료
숫 자 음	일, 이, 삼, 사, 오, 륙, 칠, 팔, 구, 영, 공, 륙

차량제어명령어는 15,20섹션별로 구별하여 실험한 결과 20섹션이 조금 더 나은 결과를 얻었기에 본 논문에서는 20섹션으로 모델을 구성하였다.



【그림 4】 모델 구성도

III. 음성인식 시스템 구현

III-1. 시스템 개발 환경

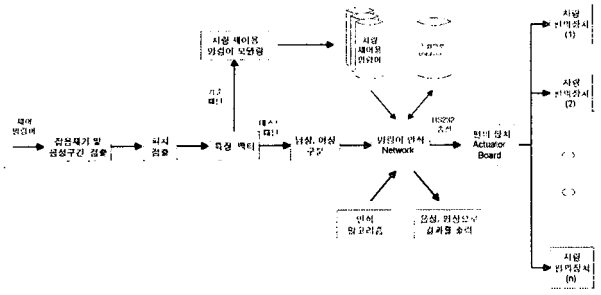
음성입력은 핀-타입 전방향성 콘덴서 마이크를 통해 이루어지며, 11.025khz 샘플링 주파수로 이산화되어 16bits로 양자화 된다. 이 과정은 일반적인 PC용 사운드 카드를 통해서 이루어지며 노트북PC상에서 비주얼 C++를 이용하여 음성신호처리 및 실험은 시내도로와 내부순환도로에서 1500cc 소형차(안반떼)를 사용하였다.

【표 1】은 본 논문에서 사용하는 제어명령어와 Voice Dialing을 위한 숫자음 30단어를 나타내고 있으며, 이들 단어는 주행중인 차량에서 화자와 마이크 거리를 30cm정도로 두고 데이터를 실험하였다.

【그림 5】 은 차량용 음성인식 시스템의 구성도를 보인다. 먼저 주행중인 자동차에서 사용자가 발성한 제어명령어는 Band Pass Filtering 알고리즘을 이용하여 잡음제거 처리 후 제어명령어 즉 음성구간을 검출한다.

검출한 제어명령어는 PLP계수[6]를 사용하여 특징벡터를 구하고 구해진 특징벡터 값을 이용하여 기준 패턴(Reference Pattern)과 테스트 패턴(Test Pattern)처리를 한다.

이렇게 인식 알고리즘(OSDP)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.



【그림 5】 차량 음성인식 장치 구성도

III-2. 인식실험 및 결과

잡음에 강인한 PLP가 다른 특징 파라미터보다 높은 인식율을 보이므로 본 논문에서는 PLP 13차를 사용하였다.

그리고, 모델은 남자, 여자 각각 5명이 10회 발성한 데이터로 사용하여 서울 시내 및 내부순환도로에서 실험하였다.

【표 2】 인식실험 결과

모델환경	화자	인식률(종속)	인식률(독립)
idle-40	남성	98.7	98
	여성	98	97
40-80	남성	97	96
	여성	95	94.4
80-100	남성	95	90.7
	여성	93	80.8

【표 2】 는 인식실험결과를 화자종속과 독립으로 구분한 결과표이다.

【표 3】 남성, 여성 모델 비교실험

모델환경	모델상태	인식률
idle-40km	남성→여성	68
	남성→남성	98.7
	여성→남성	67.9
	여성→남성	98
40-80km	남성→여성	69.3
	남성→남성	97
	여성→남성	72
	여성→여성	95
80-100km	남성→여성	41.6
	남성→남성	95
	여성→남성	48.2
	여성→여성	93

【표 3】 은 남성, 여성을 속도별로 피치검출한 결과인다. 【표 3】 에서 보듯이 남성은 남성모델로 여성은 여성모델로 인식 실험을 하여야 높은 인식률을 보인다.

IV. 결 론

본 논문은 시내주행중인 자동차 환경에서 운전자의 안전 및 편의성을 위하여, 음성인식 기술을 이용한 각종 차량 편의장치를 제어하는 실험으로 Band Pass Filter를 통해 잡음을 제거하고, 영교차율과 단구간 에너지를 이용하여 자동 음성구간 검출을 구현하였다. 영교차율과 단구간 에너지의 기준값은 1.5초마다 잡음레벨에 의해서 자동으로 보정된다. VQ는 DMS 모델을 사용하였고, 잡음환경에 강인성을 갖도록 하기 위해 기본적인 배경잡음과 주행중인 차량의 배경잡음을 적절하게 사용하여 속별로 기준모델을 구성하였다.

그리고, 화자독립과 화자종속 실험 결과 독립실험 40-80km일 때 남자 96%, 여성 94.4%의 인식률을 얻었으며, 종속일 때 남자는 97%, 여자는 95%의 높은 인식률을 얻었다.

향후 속도가 올라가므로 써 다소 인식률이 떨어지는 이유는 정확한 끝점검출을 찾지 못하거나, 예외적인 잡음 때문에 인식률이 다소 떨어지고 있으므로, 끝점검출에 대해서 계속 연구할 예정이다.

[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 Recogniton in the Car", ICASSP91, Vol.2, pp925 ~ 928, May, 14-17

[5] 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

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