성별 구별방법에 의한 자동차 내 음성 인식 성능 향상

(Improving the Performance of a Speech Recognition System in a Vehicle by Distinguishing Male/Female Voice)

양 진 우 † 김 순 협 ††
(Jin Woo Yang) (Soon Hyob Kim)

요 약 본 논문은 주행중인 자동차 환경에서 운전자의 안전성 및 편의성의 동시 확보를 위하여, 보조적인 스위치 조작 없이 상시 음성의 입, 출력이 가능한 시스템을 제안하였다. 이때 잡음에 강인한 threshold 값을 구하기 위하여, 1.5초마다 기준 에너지와 영 교차율을 변경하였으며, 대역 통과 여과기를 이용하여 1차, 2차로 나누어 실시간 상태에서 자동으로, 정확하게 끝점 검출을 처리하였다. 또한 남성, 여성을 피치검출로 구분하여 모델을 선택하게 하였고, 주행중인 자동차 속도에 따라 가장 적합한 모델을 사용하기 위하여 Idle-40km, 40-80km, 80-100km로 구분하여 남성, 여성 모델을 각각 구분하여 인식할 수 있게 하였다. 그리고, 음성의 특징 벡터와 인식 알고리즘은 PLP 13차와 OSDP(One-Stage Dynamic Programming)을 사용하였다.

본 실험은 서울시내 도로 및 내부 순환도로에서 각각 속도별로 구분하여 화자독립 인식 실험을 한 결과 40-80km 상태에서 남자는 96.8%, 여자는 95.1%, 80-100km 상태에서는 남자 91.6%, 여자는 90.6%의 인식결과를 얻을 수 있었고, 화자종속 인식실험 결과 40-80km 상태에서 남자는 98%, 여자는 96%, 80-100km 상태에서 남자는 96%, 여자는 94%의 높은 인식률을 얻었으므로, system의 유효성을 입증하였다.

Abstract This thesis discusses the design and implementation of speech recognition system for vehicle. It can bring benefits such as safety and convenience, because it doesn't require any manual operation. The threshold of energy and zero crossing rate are set to 1.5 seconds to process in noisy car environment. In addition, male and female voices are detectable by pitch. Because car environment varies depending on speed, Idling noise at 40km/h, 40km/h-80km/h, and 80-100km/h models are constructed. PLP 13th and OSDP(One Stage Dynamic Programming) are used.

The experiment was taken running along urban roads. It showed that the recognition rate for male was 96.8% and for female was 95.1% in speaker independent. At the speed of 80-100km/h, the result was 91.6% for male and 90.6% for female. In speaker dependent, male showed 98% and for male 96% at 40-80km/h and 80-100km/h, respectively. Female produced 96% and 94%.

1. 서 론

80년대의 음성인식 기술은 사무실의 조용한 환경에서 진행되어 왔으나, 90년대에서는 실생활에 밀접한 연구를 진행하고 있다. 특히 국내에서는 전자통신 연구원(ETRI)의 자동통역 시스템, 한국통신의 증권정보 안내 시스템, 윈도우즈 95 환경에서의 명령어 제어, 자동차의 편의장치 제어 등이 있으며 현재 성능개선 또는 상용화 중에 있다. 21세기에는 인간과 기계간의 가장 손쉬운 의사 소통 도구중의 하나인 음성이 인간에게 더욱 편리함을 제공 할 것이다. 최근 음성인식에 대한 많은 관심을 갖고 정부 차원에서도 연구 개발 중에 있으며 여러 분야에 집목시키고 있는 실정이며, 또한 국내·외에서는

† 정 회 원 : 춘천기능대학 전자과 교수
yjw@kopo.or.kr
†† 정 회 원 : 광운대학교 컴퓨터공학과 교수
kimsh@daisy.kwangwoon.ac.kr
논문접수 : 2000년 4월 24일
심사완료 : 2000년 10월 7일

차량용 음성인식 장치 관련된 연구개발이 활발히 진행되고 있다[1][2][3]. 본 논문에서는 기존의 음성인식 실험에서 남성화자에 대해서만 실험을 하여왔으나, 요즘들어 음성인식을 이용한 상품들이 상용화되면서 여성에 대한 관심이 한 부분을 차지하면서 여성화자 실험 또한 중요한 관건으로 대두되고 있다. 따라서, 주행중인 자동차 환경에서 또한 여성 화자 실험이 중요하다. 그렇지만, 남성 음성은 남성 모델로 여성 음성은 여성 모델로 각각 실험을 하였을 때 인식률이 높다. 하지만, 기계가 남성인지, 여성인지 구분을 하지 못하기에 본 논문에서는 음성이 입력되면 음성 인식의 전처리 부분에서 피치검출을 하여 남성의 음성인지 여성의 음성인지를 구분한 뒤에 각각의 모델을 사용하여 인식을 하게 한다.

그리고, 현재 자동차 개발의 추세를 보면, 자동차 기능은 기본 주행 기능 이외에 다양한 편의기능을 제공하게 됨에 따라 조작 버튼의 수도 상대적으로 증가되고 있는 실정이다. 이 경우 차량의 운전과 동시에 각종 장치를 조작해야 함으로써, 운전에 대한 주의 및 집중력을 저하시켜 운전자의 스트레스가 증가함은 물론, 안전운전에 상당한 악영향을 미칠 수 있다. 따라서 차량의 각종 편의장치의 조작을 음성으로 대신하게 된다면 주행시 운전자의 편의성과 안전성을 높일 수 있게 된다. 또한, 우리 나라 고속도로 사고의 사망률 중 약 19%가 카폰, 휴대폰 사용 중에 발생한다고 한다. 이는 전화번호를 누를 때 인지도의 약화와 시야 장해, 그리고 자동차의 제어시간 지연 등이 사고의 원인이라고 할 수 있다. 이러한 문제점을 해결하기 위하여 음성으로 전화를 걸 수 있도록 하는 기능은 운전자의 사고를 예방할 수 있을 뿐 아니라 사용자에게 편리성을 제공하게 된다.

그러나, 주행중인 자동차 환경에서 음성 인식을 수행하는 것은 매우 어려운 일이고, 또한 많은 문제점들이 존재한다. 먼저 주행중인 자동차 환경에서의 음성인식 장치는 보조적인 스위치 도움없이 실시간 동작이 가능해야 하며, 어느 누구나 사용가능 하면서, 안전성을 위하여 높은 인식률이 요구된다. 그리고 자동차의 엔진소리, 도로상태 등 가변적인 잡음환경에 강인한 음성인식 알고리즘의 개발이 요구된다.

본 논문의 최종목표는 주행중인 차량에서 사람이 자연스럽게 차량 제어 명령어를 말하면 자동차가 인식하여 알아듣도록 시스템을 구성하는데 있다. 그리고 주행중인 자동차에서 남성과 여성을 피치로 구분하여 인식성능을 개선하는데 다음과 같이 구성하고 있다. 1장 서론에 이어, 2장에서는 피치검출을 이용한 모델 설계와 데이타베이스 구성, 그리고 3장에서는 음성인식 시스템 구현을 바탕으로 주행중인 자동차 환경에서의 실험결과를 논하고, 끝으로 4장에서는 결론과 향후 연구과제에 대해 논의한다.

2. 피치검출을 이용한 모델 설계와 데이타베이스 구성

2.1 음성구간 자동 검출

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

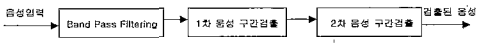

그림 1 음성구간 검출 전체 블록도

본 논문에서는 [그림 1]처럼 주행중인 차량에서 실시간으로 입력된 음성을 검출하기 위하여 대역 통과 여과를 처리하여 잡음을 제거한 후에 1차로 음성 구간과 잡음 구간을 포함한 음성을 구하고, 그리고 다시 2차로 음성 구간만을 따로 구한다.

2.2 대역 통과 여과기

2.2.1 주행 중 예외적인 잡음처리

[그림 2]는 주행중인 자동차 환경에서 끝점 검출을 하기 이전에 도로상태, 주위배경 잡음 등 예외적인 잡음이 항상 존재하고 있다. 본 논문에서는 이러한 상황 등을 고려하여 예외적인 잡음처리를 하였다. 실시간으로 동작하기 때문에 항상 마이크로폰으로 입력이 들어온다. 이때 입력된 데이타는 음성인지 잡음인지 모르지만, 프레임별로 버퍼에 저장한다. 본 논문에서는 9프레임을 1버퍼에 저장한다. 그리고 1프레임은 128 샘플로 시간 축

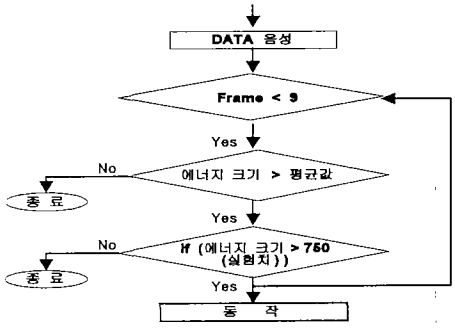

그림 2 주행중 예외적인 잡음처리

상에서 절대값을 취하여 계산한다. 이때, 270(실험치=샘플값)을 기준으로 750(실험치)샘플 이상이면 음성으로 간주하고, 750 샘플 이하이면 예외적인 잡음으로 처리한다.

1버퍼 = 9 프레임 = 1152 샘플, 1프레임 = 128 샘플

2.2.2 대역 통과 여과기 처리

사람의 음성특성은 보통 400Hz에서 4000Hz까지 분포하지만 주행중인 차량환경에서의 잡음은 0Hz에서 1600Hz까지 발생한다. 이는 차량속도 및 도로환경에 크게 영향을 받는 것으로 인식률에도 크게 영향을 끼친다. [표 1]은 자동차의 주행속도에 따른 잡음특성을 보인 것이다.

표 1 주행속도에 따른 차량잡음

주행속도	Idle 상태	40km/h 이내	80km/h 이내	100km/h 전후
잡음구간(Hz)	0 ~ 500	0 ~ 950	0 ~ 1300	0 ~ 1600

본 논문에서는 PC용 사운드 카드를 이용하여 입력된 데이타를 1400Hz에서 4200Hz 정도로 대역을 제한하며, 식 (1)은 주파수 응답을 만족하기 위한 조건이다.

$$H_r(\frac{2\pi k}{16}) = \begin{cases} 0 & k=0,\ 1,\ 7 \\ 0.456 & k=2,\ 6 \\ 1 & k=3,\ 4,\ 5 \end{cases} \quad (1)$$

그리고 [그림 3]은 16차 선형위상 유한 지연 임펄스 응답 필터의 주파수 응답을 보이며, 이에 대한 차수 값은 [표 2]와 같다.

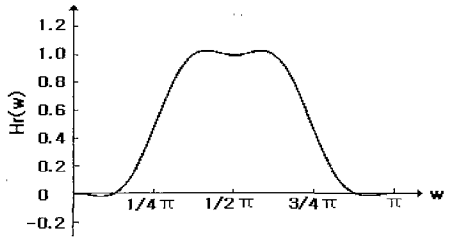

그림 3 16차 선형위상 FIR 필터의 주파수 응답

표 2 밴드패스 필터 차수 값

필터 계수	필터 계수 값	필터 계수	필터 계수 값
h(0)	0.01051756	h(8)	0.3362424
h(1)	0.02774795	h(9)	-0.266221
h(2)	0.04067173	h(10)	-0.155752
h(3)	-0.02057317	h(11)	0.04840164
h(4)	0.04840164	h(12)	-0.02057317
h(5)	-0.155752	h(13)	0.04067173
h(6)	-0.266221	h(14)	0.02774795
h(7)	0.3362424	h(15)	0.01051756

[그림 3]은 대역 통과 여과기의 스펙트럼 형태로써, 이와 같은 필터계수와 배경잡음이 섞인 음성 신호의 필터링은 다음과 같다.

$$y(n) = \sum_{k=1}^{N} x(n-k)h(k) \quad (2)$$

식 (2)에서 $x(n)$은 배경 잡음이 섞인 음성 신호이고, $h(k)$는 필터 계수이다. 이때 $y(n)$은 필터링한 출력 값이고, N은 16차를 가르킨다.

2.3 음성구간 검출

음성 구간과 묵음 구간을 분리해 내는데 가장 널리 이용된 방법은 영 교차율과 단구간 에너지이다[4]. 하지만 주행중인 자동차의 배경 잡음에서는 영 교차율과 단구간 에너지만으로는 끝점 검출을 한다는 것이 매우 어렵다. 본 논문에서는 영 교차율과 단구간 에너지를 이용하기 전에 식(2)을 이용하여 잡음을 제거한다.

$$ZCR(i) = \sum_{n=0}^{N-1} sgn(y_{n+(i-1)*N}\ y_{n+1+(i-1)*N}) \quad (3)$$

$$E(i) = \sum_{n=1}^{N} |y((i-1)*N+n)| \quad (4)$$

식(3)에서 i는 프레임의 번호이고 ZCR(i)는 i번째 프레임의 영 교차율 수이며, N은 샘플수 128을 가르킨다. 식(4)에서 $E(i)$는 i번째 프레임 에너지 값의 합을 의미한다. 본 논문에서는 음의 크기를 감소시킨 잡음과 음성구간을 좀 더 정확하게 구별하기 위하여 식(3), (4)에 제곱을 한다.

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

$$\overline{E} = E(i)^2 \quad (6)$$

$$\overline{E} < E(i)^2 \times 3.5 \quad (7)$$

$$\overline{ZCR} < ZCR(i)^2 \quad (8)$$

기준 threshold는 잡음에 적응시키기 위하여 1.5초마다 재조정하고, 식(7), (8)의 $\overline{E}$, $\overline{ZCR}$은 식(5), (6)에서 구한 threshold의 에너지 값과 영 교차율이며, E, ZCR은 실시간으로 입력되는 잡음 또는 잡음 섞인 음성구간의 에너지 값과 영 교차율을 가르킨다. 음성구간의 시작부분을 구하기 위해서는 식(7), (8)을 이용한다. 식(7), (8)을 동시에 연속적으로 4프레임이상 만족하면 음성구간으로 간주한다. 만약 이 조건을 만족하지 않으면 잡음구간으로 처리한다.

음성의 끝 부분 확인은 음성이 끝난 이후에 0.5~0.7초 동안에 잡음이 계속되면, 즉 식(7), (8)을 만족하지 않으면 음성입력이 끝난 것으로 처리한다.

2.4 피치 검출[5]

1939년 R. Dudley에 의해 보코더의 이론이 제안된 후부터 유성음의 여기신호 주파수인 피치검색에 관한

연구도 주된 처리영역에 따라 시간영역, 주파수영역, 시간-주파수 혼성영역으로 분리해서 다루어지고 있다. 피치의 범위는 보통 2.5에서 25msec로 알려져 있고 음성을 8㎑로 표본화하여도 그 범위가 20~200 표본사이에 나타나기 때문에 시간 영역에서의 검출은 분해능이 높다는 장점이 있다. 보통 음성의 분석은 프레임 단위로 처리되지만 파열음의 경우, 음소의 변화가 한 프레임 안에서도 일어날 수 있다. 따라서 시간 영역에서의 피치 측정은 음소 변화에 따른 특성을 검출할 수도 있다. 그렇지만, 음소의 종류나 결합위치에 따라 음소 파형의 외곽 선이 변화하는데, 이 변화를 측정하기 어렵기 때문에 이러한 파형 부분에서 피치검출이 간단하고 정확한 추출이 되지만 스펙트럼을 계산하고 자기상관계수를 구하는 과정이 프레임을 구성하는 표본 전 구간에 대해 계속 적용되기 때문에 처리시간이 길어진다. 실시간으로 실행이 가능하게 하기 위해선 처리시간의 속도가 중요한 의미를 가진다

2.4.1 자기상관 함수

자기상관 함수는 에너지 스펙트럼의 역 변환인 시간 신호를 말한다. 자기상관 함수는 신호의 고조파와 포만트 진폭, 주기 정보를 유지하고, 위상정보는 무시한다. 이 자기상관 함수는 피치검출, 유/무성음 결정 그리고 선형예측에 사용된다.

자기상관 함수는 다음과 같은 상호상관 함수의 특수한 경우이다.

$$\phi(k)=\sum_{m=-\infty}^{\infty}x(m)y(m+k) \tag{9}$$

상관함수는 신호 $x(n)$과 $y(n)$사이에서 시간 지연으로 유사 도를 측정한다.

신호표본과 지연된 다른 신호 표본과의 곱의 합에 의해서 두 신호의 유사 도를 측정한다. 만약 두 신호가 비슷한 파형이라면 상호 상관 값은 크게 나타난다.

$x(n)$과 $y(n)$의 신호가 같을 때, 식 (9)은 자기상관 함수의 정의식인 식(10)으로 쓸 수 있다.

$$\Phi(k)=\sum_{m=-\infty}^{\infty}x(m)x(m+k) \tag{10}$$

선형 예측의 경우에 $\Phi(k)$에서 최대 값을 갖는 k의 표본 간격으로 피치를 검출한다.

자기상관 함수는 지연이 크면 클수록 자기신호의 곱이 적게 된다. 이러한 방법은 곱과 합을 반복함으로써 수행하는 시간이 많이 걸린다. 따라서 이러한 방법의 변형인 AMDF 방법을 사용하여 피치를 추출하고자 한다.

2.4.2 AMDF(Average Magnitude Difference Function)

자기상관 함수에서 $x(n)$과 $x(n+k)$곱 대신에 다음과 같이 절대값으로 정의된다.

$$AMDF(K)=\sum_{m=0}^{N-1}|x(m)-x(m-k)| \tag{11}$$

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

자기상관 함수 $\Phi(k)$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 피치를 추출해낸 다음 기본주파수로 변환시킨다. 피치와 기본주파수는 같은 뜻으로 쓰이는데 피치는 시간 축에서 음성 파형의 주기를 찾아내는 것을 말하고, 이러한 피치를 주파수 축에서 해석한 것을 기본주파수라고 한다. $f=\frac{1}{T}$ (여기서 f=주파수,T=시간주기)이므로 구해진 피치 값을 시간주기로 바꿔주면

1초 : 11025 sample = x초 : 피치 샘플이므로

$x초=\frac{pitch}{샘플링주파수}$ 가 된다.

(본 논문에서 샘플링 주파수는 11025Hz 사용)

따라서, $f=\frac{1}{\frac{pitch}{11025}}$ 가 된다. 결론적으로 $f=\frac{11025}{pitch}$ 이 된다.

입력된 음성을 가지고 피치를 구한다. 앞에서 설명하였듯이 한 프레임에 하나의 피치가 나온다. 만약, "비상등"이 20 프레임이었다면 20개의 피치 값이 나오는 것이다.

예를 들어 86, 120, 109, 98 데이타 값이 있으면 가장 큰 값과 가장 작은 값을 제외한 109, 98 값의 평균 103.5 값을 사용한다.

다른 예로 0, 0, 112, 109, 120, 139, 521, 521, 182, 182, 0, 0 데이타 값이 있으면 일반적으로 남성은 350Hz 보다 큰 값, 여성은 50Hz 보다 작은 값은 아주 드물다. 그렇기 때문에 50 - 350Hz로 제한을 두었다. 이때 0과 521Hz는 제거를 한다. 그리고, 112, 109, 120, 139, 182, 182의 평균 값 146.57 값을 구할 수 있다. 그 중에서 최소 값 109, 최대 값 182를 제외한 평균 값 123.67 값을 사용한다. 그 이유는 안정된 값을 사용하기 위하여 위와 같은 제한을 두었다.

[그림 4]의 AMDF법은 자기상관 함수법에 수행하는 곱 연산을 절대값과 차분으로 대신하기 때문에 상대적

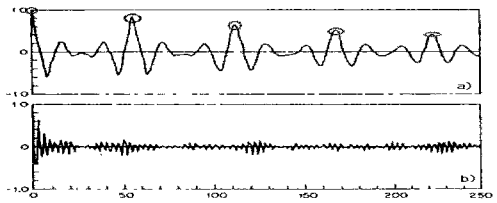

그림 4 유/무성음에 대한 자기상관 함수법 (a)유성음 b)무성음)

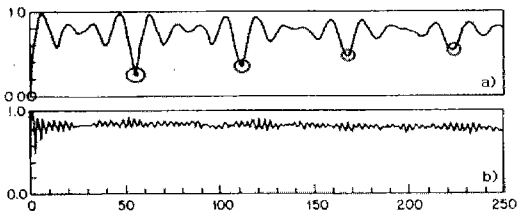

그림 5 유/무성음에 대한 AMDF법 (a)유성음 b)무성음)

으로 빠르다는 장점을 가지고 있다. 이런 이유는 실시간에 많이 적용한다. 그리고 자기상관 함수는 피치주기 배수가 최대 값을 이루지만, AMDF법에서는 피치주기 배수가 최소 값을 갖는다.

[그림 4, 5]에서 첫 번째 최소 값이 음성 프레임의 기본주파수 즉, 피치가 된다. 본 논문에서는 주행중인 자동차 환경에서 남성, 여성 각각 5명이 10회 발음한 데이타를 가지고 실험을 하였다.

[그림 6]은 1500cc 소형 승합차에서 차량 제어 명령어를 발성했을 때 피치 값의 평균과 표준편차로 나타낸 정규분포도이다. 일반적으로 피치범위는 남성의 경우, 가능한 피치범위가 50 ~ 250Hz 범위에 존재하고, 여성의 경우에는 120 ~ 500Hz 범위에 존재한다.

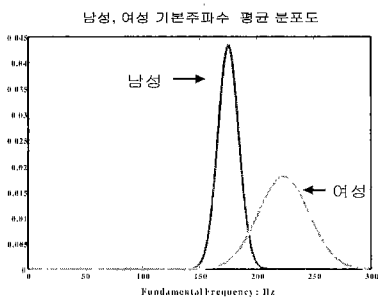

그림 6 자동차 환경에서 남성, 여성기본주파수 평균 분포도

[그림 6]은 자동차 환경에서 차량 제어 명령어를 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.

[표 3]은 남성, 여성 속도에 따라 피치분포에 따른 정규 분포 표이다.

표 3 남성, 여성 속도에 따라 피치분포에 따른 정규 분포표

환경	평균	최소값	최대값	표준편차
남자 Idle	157	113	208	16.8
남자 40-80km	179	116	230	5.4
남자 80-100km	187	136	263	5.4
여자 Idle	228	153	294	26.6
여자 40-80km	224	128	298	20.2
여자 80-100km	216	105	298	19.8
남자	174.3	113	263	9.2
여자	222.7	105	298	22.2

2.5 주행중인 자동차 환경에 적합한 모델 설계

음성모델은 인식하고자 하는 어휘들간의 변별력이 크도록 구성되어야 하며, 특히 이 변별력은 음성인식 시스템의 동작환경, 즉 잡음환경, 화자의 발음 습관의 변화에도 유지되어야 한다. 결국 음성모델은 언어정보에 대해 변별력이 커야 하며 음성 정보에 대해서는 민감하지 않아야 한다. 시스템이 가변적인 동작환경에 강인하기 위해서는 음성모델 구성을 위한 음성의 취득환경, 언어정보를 특징짓는 특징 치의 선택이 중요하다. 예를 들어 실험실 환경의 깨끗한 음성으로 구성된 음성모델은 100km/h의 주행환경에서는 낮은 인식률을 보이게 된다. 또한 모델구성에 참여하는 화자의 발음 횟수가 증가할수록 발성습관의 변화에 대해서는 강인하겠지만 어휘간

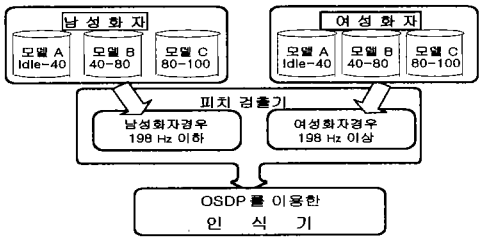

그림 7 주행중인 자동차 환경에 적합한 모델 구조

의 변별력은 감소시킬 수 있다. 결국 음성모델을 구성하는데 있어서 잡음요소를 첨가함으로써 잡음에 대한 강인성을 높일 수 있으며, 화자의 발음횟수를 적절히 선택함으로써 화자의 발성 습관의 변화에 대처할 수 있다.

본 논문은 어휘간의 변별력을 높이기 위하여 [그림 7]처럼 남성화자, 여성화자를 구분하여 모델 A, B, C 3개로 나누고, 주행중인 자동차 환경에서 속도별로 남성, 여성을 구분하여 가변적인 잡음에 강인하기 위하여 Idle-40km, 40-80km, 80-100km 환경으로 실험하였다.

[그림 7]에서 모델 A는 Idle-40km, 모델 B는 40-80km, 모델 C는 80-100km 환경이다.

2.6 음성 데이타베이스 구성

2.6.1 구성 명령어

주행중인 자동차 환경에서의 음성인식은 사람의 목소리 즉 음성을 이용하여 차량의 각종 편의장치를 제어하기 위한 것이므로 [표 4]처럼 차량을 제어하기 위한 18개의 고립단어와, 전화번호 등록 및 제어기능을 위한 숫자음 고립단어 12개로 구성한다. 이중에서 통화시작과 통화 종료는 주행중인 차량에서 운전자가 카폰, 핸드폰을 사용할 때 사용한다.

표 4 차량 편의장치 제어용 데이타베이스

비상등	와이퍼	실내등	오디오	에어컨
히터	정지	다음 채널	이전 채널	라디오 스캔
소리 크게	소리 작게	온도 올려	온도 내려	창문 올려
창문 내려	통화 시작	통화 종료	일	이
삼	사	오	육	칠
팔	구	영	공	륙

2.6.2 데이타베이스 취득 및 구성방법

잡음환경에 강인한 차량제어용 명령어 모델생성과 속도, 성별로 구분하여 Idle-40km, 40-80km/h, 80-100 km/h 상태의 음성 데이타를 취득한다. 주행도로는 서울시내 도로, 내부순환 도로 및 아스팔트 고속도로를 이용한다. 또한 도로의 주변환경(아스팔트, 시멘트 도로, 옆 차선의 차량 지나가는 소리, 도로상태, 차량엔진 등)은 모두 수용한 상태에서 데이타를 취득한다.

3. 음성인식 시스템 구현

3.1 시스템 개발 환경

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

[표 4]는 본 논문에서 사용하는 제어명령어와 음성 다이얼링을 위한 숫자음 30단어를 나타내고 있으며, 이들 단어는 주행중인 차량에서 화자와 마이크 거리를 30 ㎝정도로 두고 데이타를 실험하였다.

[그림 8]은 차량용 음성인식 시스템의 구성도를 보인다. 먼저 주행중인 자동차에서 사용자가 발성한 제어명령어는 대역 통과 여과기 알고리즘을 이용하여 잡음제거 처리 후 제어 명령어 즉 음성구간을 검출한다. 검출한 제어명령어는 PLP 계수[6]를 사용하여 특징벡터를 구하고 구해진 특징벡터 값을 이용하여 기준 패턴과 테스트 패턴처리를 한다.

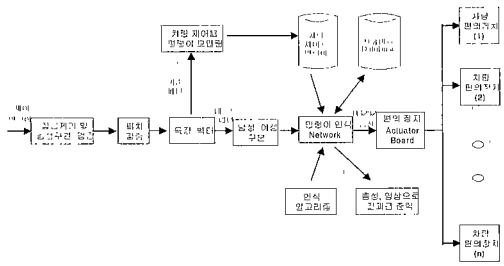

그림 8 차량 음성인식 장치 구성도

이렇게 인식 알고리즘(OSDP)[7]을 사용하여 구해진 명령어 결과는 스피커를 통하여 음성으로 출력하고, RS232 케이블을 이용하여 편의장치인 액튜에이터 보드로 전송한다. 그리고 편의장치인 액튜에이터 보드는 전송된 결과를 이용하여 차량 편의장치를 기동한다.

3.2 인식실험 및 결과

잡음에 강인한 PLP가 다른 특징 파라미터보다 높은 인식률을 보이므로 본 논문에서는 PLP 13차를 사용하였다. 그리고, 모델은 남자, 여자 각각 5명이 10회 발성한 데이타로 사용하여 서울시내 및 내부 순환도로에서 실험하였다.

3.3 주행중인 차량환경에서의 모델선정

DMS 모델은 구간을 단어의 특성에 따라 동적 구간으로 나누고, 이 동적 구간이 각 구간의 대표 특징 벡터와 지속시간 정보를 가지도록 만들어야 하기 때문에 모델을 작성하는데 좀 더 복잡한 절차가 필요하다. DMS 모델을 작성하기 위해서는 크게 두 단계로 분리되어서 수행되어진다. 먼저, 선행하는 단계는 구간을 동적으로 분할하는 구간 구분화 작업이고, 다음 단계는 구분된 구간들에 대해 각 구간의 대표 특징 벡터와 지속시간 정

보를 구하는 단계이다.

본 논문에서는 수행 속도 및 인식 성능 향상을 위하여 DMS 모델의 한 섹션 당 일정한 지속시간을 가진 음성 신호를 할당하여 구간을 나누는 개선된 DMS 모델을 사용하고, 제안하고 있는 신호의 지속시간을 고려한 가변 섹션 모델을 생성하기 위해서는 취득된 각 인식 단어에 대한 데이타들의 평균 지속 시간에 대한 정보와 지속 시간과 섹션의 수에 따른 인식률의 변화를 알아야 한다.

본 실험을 통하여 같은 조건에서 섹션의 수를 변경하여 단어에 대한 지속 시간에 따른 인식률의 변화를 확인하고자 한다. [표 5]는 차량환경에 강인한 모델을 선정하기 위하여 Idle-40km, 40-80km, 80-100km 속도별로 구분하여 남·여 각각 실험하였다.

[표 5]의 인식률 결과표는 차량 제어명령어를 각각 20, 15 섹션으로 구분하여 실험한 결과이다.

[표 5]에 보았듯이 20 섹션 일 때가 다소 차이가 있지만 더 좋은 결과를 보였다. 그래서 본 논문에서는 차량 제어명령어는 20 섹션으로 만든 모델을 사용한다.

[표 6]은 20 섹션으로 만든 모델로 독립 실험한 결과표이다.

표 5 인식률 결과표 (화자종속)

모델환경	모델상태		인식률(%)	개수(오인식)	기 타
	화자	섹션			
Idle-40km	남성	20	98.7	900(12)	남성 Idle
		15	98 7	900(12)	남성 Idle
	여성	20	98	900(18)	여성 Idle
		15	98	900(18)	여성 Idle
40-80km	남성	20	98	540(11)	남성 40km
		15	97.8	540(12)	남성 40km
	여성	20	96	540(22)	여성 Idle+40(1+3)
		15	96	540(22)	여성 Idle+40(1+3), 여성 40km
80-100km	남성	20	96	540(22)	남성 40+80km
		15	95.3	540(25)	남성 80km, 남성 40+80km
	여성	20	94	540(32)	여성 80km
		15	94	540(32)	여성 80km

표 6 인식률 결과표 (화자독립)

모델환경	모델상태		인식률(%)	개수(오인식)	기 타
	화자	섹션			
Idle-40km	남성	20	98	900(18)	남성 Idle
	여성	20	97	1200(36)	여성 Idle
40-80km	남성	20	96.8	450(14)	남성 40km
	여성	20	95.1	450(22)	여성 Idle+40(1+3)
80-100km	남성	20	91.6	450(38)	남성 40+80km
	여성	20	90.6	450(42)	여성 80km

[표 5, 6] 실험은 남성 모델로 남성 실험을 하였을 때나 여성모델로 여성 실험을 하였을 때는 별 문제가 없다. 그렇지만, 남성모델로 여성을 실험하였을 때나 반대로 여성모델로 남성을 실험하였을 때는 인식률에 차이 심하다 그래서 본 논문에서는 자동차 환경에서 남·여성을 구별하고자 피치검출법의 한 방법인 AMDF (Average Magnitude Difference Function) 방법을 사용하였다.

AMDF법은 자기상관 함수 법에서 수행하는 곱 연산을 절대값과 차분으로 대신 하기 때문에 상대적으로 빠르다는 장점을 가지고 있다. 이러한 이유로 실시간에 많이 적용하고 있다.

3.4 주행중인 차량환경에서의 성별이 구별되어야 하는 이유

본 논문에서 제안한 방법으로 피치검출을 하여야 하는 이유는 [표 7]에서 보듯이 남성모델로 여성을 실험하였을 경우 오 인식이 높다. 또한 여성모델로 남성을 실험하였을 경우도 마찬가지다.

그래서, 피치검출을 하여 남성은 남성모델로 여성은 여성모델로 인식 실험을 하여야 높은 인식률을 보인다.

[표 8]은 피치검출을 하였을 때 남성·여성을 구분한 결과를 백분율로 나타낸 결과표이다.

차량속도가 올라가므로 써 남, 여 피치검출에 약간의 오 인식이 발생한다.

그렇지만, 여성화자의 목소리가 남성처럼 두꺼운 사람 즉 중·저음 쪽에 가까운 화자는 남성모델에서 인식하는데 더 나은 결과를 보이고 있다.

남성화자 또한 여성화자의 목소리처럼 가는 목소리인 고음 쪽에 가까운 화자는 여성모델에서 인식하는데 더 나은 결과를 보이고 있다.

표 7 남성·여성 모델 비교실험

남성, 여성 모델 비교실험				
모델환경	모델 상태	인식률	개수(오인식)	인식률 차이
Idle-40km	남성 → 여성	68	900(288)	32
	남성 → 남성	100	900(0)	
	여성 → 남성	67.9	900(289)	31.8
	여성 → 여성	99.7	900(3)	
40-80km	남성 → 여성	69.3	540(166)	30.7
	남성 → 남성	100	540(0)	
	여성 → 남성	72	450(126)	26.7
	여성 → 여성	98.7	450(6)	
80-100km	남성 → 여성	41.6	450(263)	58.2
	남성 → 남성	99.8	450(1)	
	여성 → 남성	48.2	450(233)	51.4
	여성 → 여성	99.6	450(2)	

표 8 피치 검출 결과표

환경		화자	인식(개수)
Idle-40km	남자	남자	750(747)
		여자	750(3)
	여자	남자	750(48)
		여자	750(702)
40-80km	남자	남자	450(425)
		여자	450(25)
	여자	남자	540(114)
		여자	540(486)
80-100km	남자	남자	450(372)
		여자	450(78)
	여자	남자	600(244)
		여자	600(356)

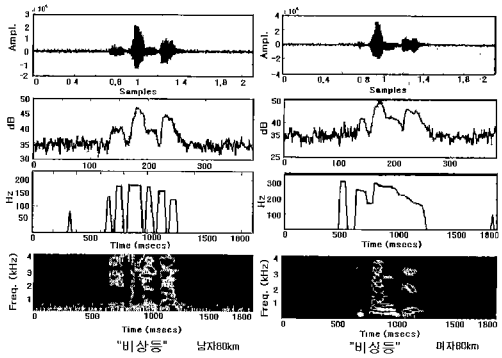

그림 9 남성·여성화자의 파형, 에너지, 피치, 스펙트로그램

[그림 9]는 80km 속도에서 남성·여성화자가 "비상등"이라는 단어를 발음하였을 때 파형, 에너지, 피치, 스펙트로그램을 보여주고 있고 남성과 여성은 피치 주기가 다를 뿐만 아니라 여성의 목소리가 고음에 더 많이 분포하고 있음을 알 수 있다.

4. 결 론

주행중인 자동차 환경에서의 편의성 및 안전성의 동시 확보를 위한 음성인식 시스템을 구현하기 위해서는 보조적 스위치의 조작 없이 상시 음성의 입력이 가능해야 하며, 자동으로 음성구간 검출이 이루어져야 한다. 그리고 자동차의 도로주행 환경의 특성상 가변적인 변화는 환경에 강인해야 한다. 본 논문에서는 밴드패스 필터를 통해 잡음을 제거하고, 영 교차율과 단구간 에너지를 이용하여 자동 음성구간 검출을 효과적으로 구현하였다. 영 교차율과 단구간 에너지의 기준 값은 1.5 초마다 잡음레벨에 의해서 자동으로 보정된다. VQ는 DMS [8]모델을 사용하였고, 잡음환경에 강인성을 갖도록 하기 위해 기본적인 배경잡음과 주행중인 차량의 배경잡음을 적절하게 사용하여 속도별로 기준모델을 구성하였다. 또한 21세기는 남·여성이 모든 일에 함께 참여하기 때문에 한쪽에 편중해서 실험을 해서는 안 된다. 그래서 본 논문은 남·여성에 대한 실험을 하였고 인식률 결과를 보았을 때 남성모델로 남성 인식실험에는 좋은 결과를 보이지만, 남성모델로 여성 인식실험을 하였을 때 인식률에 차이가 많았음을 볼 수 있었다. 여성 또한 같은 결과를 보여주고 있다.

따라서, 본 논문에서 제안한 피치검출 방법으로 실시간에 많이 사용하는 AMDF 알고리즘을 사용하여 위 문제점을 보안하였다.

그리고, 차량속도에 따라서 가변적인 모델을 사용하여 Idle-40km, 40-80km, 80-100km에 맞게 남·여성을 구분하여 모델을 만들어 인식률을 높였다. 화자독립과 화자종속 실험결과 독립실험 40-80km일 때 남자는 96.8%, 여자는 95.1%의 인식률을 얻었으며, 종속일 때 남자는 98%, 여자는 96%의 높은 인식률을 얻었다.

본 논문에서는 시스템의 설계 및 구현 후에 실시간으로 많은 인식실험을 하였다. 인식실험 후 에러를 분석한 결과 '창문 올려' → '창문 내려', '소리 크게' → '소리 작게', '온도 올려' → '온도 내려', '칠' → '삼' 단어에서 오 인식이 많이 발생하였다. 각 단어를 분석한 결과 화자가 음성을 발성할 때의 다소 차이가 나타남을 볼 수 있었다.

그리고, 주행중인 자동차환경에서 보조적 스위치의 조작 없이 상시 음성의 입력이 가능해야 하며, 자동으로 음성구간 검출을 하여 음성인식이 수행되어야 하므로 옆 사람과의 대화, 라디오, 오디오 소리 등의 효과적인 음성 거절기능의 처리가 필요하며, 본 논문에서는 주행중인 자동차 환경에서 예외적인 잡음 (노면 턱 넘을 때 나는 소리, 지나가는 화물차 소리 등)처리를 함으로써 인식 율을 향상시킬 수 있었다. 앞으로도 더 많은 실험을 통하여 이 부분을 계속 연구해야할 것이며, 피치검출을 통하여 남성, 여성화자를 구분한 모델로 무작위로 실험을 했을 때 인식률과의 차이점을 찾아야 할 것이다. 또한 다양한 화자(연령별)실험을 계속 진행할 것이며, 고속도로 및 시멘트도로 환경에서도 더 많은 실험을 계속 진행할 것이다. 또한, 향후 속도가 올라감으로써 다소 인식률이 떨어지는 이유는 정확한 끝점검출을 찾지 못하거나, 예외적인 잡음 때문에 인식률이 다소 떨어지고 있으므로, 끝점 검출에 대해서 계속 연구할 예정이다.

참 고 문 헌

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

[2] A. Noll, "Problem of Speech Recognition in Mobile Environments," ICSLP90, Vol.2, pp.1133~1136, 1990.

[3] Chafic MOKBEL, Ge'rard CHOLLET, "An Improved Noise Compensation Algorithm for Word Recognition in the Car," ICASSP91, Vol.2, pp.925~928, May 14-17

[4] L.R. Rabiner, M. R. Sambur, "An Algorithm for Determining the Endpoints of Isoated Utterances," The Bell System Technical Journal, Vol.54, No.2, pp.297~315, Feb. 1975.

[5] L.R. Rabiner, "On the use of Autocorrelation Analysis for Pitch Detection," J. Acoust., Speech, Signal processing, Vol. ASSP-25, pp.24~33, Feb. 1977.

[6] H. Hermansky, "Perceptual Linear Predictive (PLP) Analysis of Speech," J. Acoust. Soc. Am. 87(4), pp.1738~1752, Apr. 1990.

[7] H. 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, pp.263~271 Apr. 1984.

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

[9] L.R. Rabiner, B.H. Juang, "Fundamentals of Speech Recognition," Prentice Hall, 1993.

[10] H.G. Hirsch, P. Meyer and H.W. Ruehl, "Improved Speech Recognition Using High-Pass filtering of Subband Envelopes," EUROSPEECH91, Vol.2, pp.413~416, Sep. 1991.

[11] P. Lockwood, C. Baillargeat, J.M. Gillot, J. Boudy, G. Faucon, "Noise Reduction for Speech Enhancement in Cars: Non-Linear Spectral Subtraction/ Kalman Filtering," EUROSPEECH91, pp. 83~86, Vol.1, Sep. 1991.

[12] L.R. Rabiner, M.R. Sambur, "An Algorithm for Determining the EndPoints of Isolated Utterances," The Bell System Technical Journal, Vol.54, No.2, pp.297~315, Feb. 1975.

[13] 배명진, 이상효 "디지탈 음성분석," 동영출판사, 1998.

[14] 이정기, 남동선, 양진우, 김순협 "실시간 윈도우 환경에서 DMS모델을 이용한 자동 음성 제어 시스템에 관한 연구", 한국음향학회지, 19권 3호, pp.51~56, Apr. 2000.

[15] 양진우, 김순협 "주행중인 자동차 환경에서의 음성인식 연구", 한국음향학회지, 19권 5호, pp.3~8, July 2000.

양 진 우

1982년 2월 원광대학교 전자공학과 (공학사). 1985년 2월 광운대학교대학원 전자공학과(공학석사). 1994년 8월 광운대학교대학원 전자계산기공학과 (박사과정 수료). 1982년 ~ 현재 한국음향학회 정·종신회원, 학술위원. 1996년 5월 ~ 현재 춘천기능대학 전자과 교수. 주관심분야 음성인식, 디지털신호처리, *HCI* 등.

김 순 협

1974년 울산대학교 전자공학과 (공학사). 1976년 연세대학교대학원 전자공학과 (공학석사). 1984년 연세대학교대학원 전자공학과 (공학박사). 1986년 8월 ~ 1987년 8월 Univ. of Texas at Austin, Dept. of Electronical & Computer Eng 객원교수. 1991년 1월 ~ 1993년 12월 대 검찰청 과학수사연구소 자문위원. 1998년 1월 ~ 1999년 12월 한국음향학회 회장. 1979년 3월 ~ 현재 광운대학교 컴퓨터공학과 교수. 주 관심분야는 음성인식, 디지털신호처리, 임체음향, HCI 등.