

論文93-30B-2-6

자동차 소음 환경에서 음성 인식 (Speech Recognition in the Car Noise Environment)

金元九*, 車日煥*, 尹大熙*

(Weon Goo Kim, Il Whan Cha and Dae Hee Youn)

要約

본 논문은 자동차 소음 환경에서 음성 다이얼링을 위한 화자 종속 단독음 인식 시스템의 개발에 관한 연구이다. 이러한 연구를 위하여 자동차 소음 환경에서 음성 인식 시스템의 성능을 개선할 수 있는 방법들이 제안되었고, 시속 100 km로 주행중인 자동차에서 수집한 데이터 베이스를 사용하여 평가되었다. 본 연구에서 사용한 방법들은 다음과 같다. 배경 잡음을 AR 모델링한 필터의 계수를 사용하여 역 필터링한 신호의 크기를 사용한 음성 구간 검출 방법을 수행하여 검출 오차를 감소시켰으며, 발생한 오차는 변형된 DTW 알고리즘을 사용하여 보상하였다. 한편, 선형 예측 과정에서 구할 수 있는 잔차 에너지가 양수가 되는 제약 조건을 갖는 자기 상관 차감법을 사용하여 잡음을 제거하였다. 또한 소음에 강한 거리 측정 방법을 사용하여 특징 벡터의 왜곡을 감소시켰다. 음성 인식 시스템은 범용 신호 처리 프로세서인 모토로라사의 DSP56001을 사용하여 구현되었다. 데이터 베이스는 3명의 남성 화자가 발음한 50개의 이름으로 구성된다. 인식 오차는 각 단어당 한개의 기준 패턴을 사용한 경우에 4.3%, 각 단어당 2개를 사용한 경우에 1.5%였다.

Abstract

This paper describes the development of a speaker-dependent isolated word recognizer as applied to voice dialing in a car noise environment. For this purpose, several methods to improve performance under such condition are evaluated using database collected in a small car moving at 100km/h. The main features of the recognizer are as follow : The endpoint detection error can be reduced by using the magnitude of the signal which is inverse filtered by the AR model of the background noise, and it can be compensated by using variants of the DTW algorithm. To remove the noise, an autocorrelation subtraction method is used with the constraint that residual energy obtainable by linear predictive analysis should be positive. By using the noise robust distance measure, distortion of the feature vector is minimized. The speech recognizer is implemented using the Motorola DSP56001(24-bit general purpose digital signal processor). The recognition database is composed of 50 Korean names spoken by 3 male speakers. The recognition error rate of the system is reduced to 4.3% using a single reference pattern for each word and 1.5% using 2 reference patterns for each word.

1. 서론

* 延世大學交 電子工學科

(Dept. of Elec. Eng. Yonsei Univ.)

接受日字 1992年 9月 14日

(*본연구는 삼성전자의 연구비 지원에 의하여 이루어 진것임.)

자동차용 전화가 발달함에 따라서 자동차 환경에서의 잡음 제거와 음성 인식에 대한 관심이 높아지고 있다. 이러한 음성 다이얼링 시스템은 운전자의 운전

을 방해하는 버튼식 전화기를 보다 안전한 방법으로 대체할 수 있게 한다. 그러나, 비교적 조용한 환경에서는 잘 동작하는 음성 인식 시스템의 성능은 주변 소음이 증가함에 따라서 급격히 감소한다. 소음이 패턴 매칭 방법을 사용하는 음성 인식 시스템의 성능을 저하시키는 가장 큰 원인은 음성 구간 검출 오차와 특징 벡터가 소음의 영향으로 왜곡되는 것이다. 평균 100 km/h의 속도로 고속도로를 주행하는 자동차 속에서, 자동차의 핸들 또는 햇빛 가리개에 설치한 마이크로 들어오는 입력 음성의 신호대 잡음비는 0 dB 또는 그 이하로 떨어지기도 한다. 이러한 환경에서는 음성 구간 검출도 매우 어려워진다. 또한 음성 구간 검출 오차는 음성 인식 시스템 전반에 영향을 미친다. 따라서 음성 다이얼링 시스템을 자동차 환경에서 사용하기 위해서는 소음에 대한 대책이 필요하다.

소음에 강한 음성 인식 시스템에 관한 연구는 몇가지 방향으로 진행되어 왔다.^[1, 2, 4, 10] 즉, 소음이 첨가된 신호로부터의 잡음 제거, 소음에 강한 특징 벡터나 거리 측정 방법을 사용하거나 음성 인식 시스템을 소음 환경에 적응시키는 것이다. 첫번째 방법은 실험실 환경에서 성공을 거둔 음성 인식 시스템을 소음 환경에 대하여 강하게 만드는 방법으로서 인식 시스템의 입력 전단에서 음성에 섞인 소음을 제거하는 잡음 제거 시스템을 부착하는 것이다.^[3] 두번째 방법은 음성 인식 시스템을 설계할 때부터 입력 음성에 첨가되는 소음에 강하게 제작하는 것이다.^[2] 이러한 방법은 일반적으로 사용되는 방법과는 다르게 소음에 영향을 적게 받는 특징 벡터나 거리 측정 방법을 사용하는 인식 시스템을 설계하는 것이다. 세번째 방법은 기준 패턴을 소음 환경에 적합하도록 변화시키는 것이다.^[4]

본 연구에서는 자동차 소음 환경에서 음성 인식 시스템의 구현을 위하여 소음 환경에서 발생하는 문제점을 분석하고 음성 인식 시스템의 성능을 개선할 수 있는 방법들에 관하여 연구하였다. 또한 음성 인식 시스템은 범용 DSP 칩인 모토롤라사의 DSP56001을 사용하여 구현하였으며^[5], 100 km/h로 주행 중인 자동차에서 발음한 음성 데이터를 사용하여 인식 시스템의 성능을 평가하였다.

II. 음성 인식 시스템

자동차 소음 환경에서 사용된 음성 인식 시스템의 블럭도는 그림 1과 같다. 음성 인식 시스템은 전처리 과정(pre-processing), 배경 잡음 레벨 추정(background noise level estimation), 음성 구간 검출(endpoint detection), 잡음 제거(noise

subtraction), 특징 파라메타 추출(LPC analysis), 소음에 강한 거리 측정 방법(noise robust distance measure), 변형된 패턴 매칭(modified DTW)과 결정 법칙(decision rule)으로 구성된다. 본 절에서는 소음에 대한 대책을 갖춘 음성 인식 시스템의 각 부분에 대하여 기술한다.

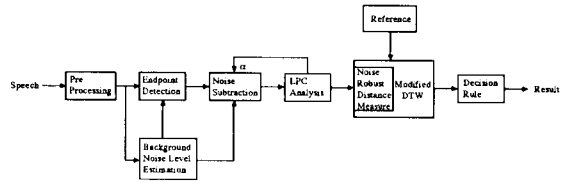


그림 1. 음성 인식 시스템

Fig. 1. Speech recognition system.

1. 전처리 과정

자동차 소음은 대부분의 에너지가 400 Hz 이하의 저주파 영역에 집중된 유색 잡음(colored noise)이므로 250 Hz의 고역 통과 필터를 사용하면 자동차 소음이 첨가된 음성 신호에 거의 영향을 미치지 않으면서 상당 부분의 소음을 제거할 수 있다.^[10] 따라서 본 연구에서는 자동차 소음을 줄이는 방법의 하나로서 자동차 소음이 섞인 음성을 고역 통과 필터를 통과시켜서 소음을 줄인 후에 음성 인식을 수행하였다.

2. 음성 구간 검출

음성 구간 검출 과정은 입력 신호로부터 음성 신호를 검출하는 과정으로서 음성 구간 검출의 결과가 전체 시스템의 인식율에 미치는 영향이 크므로 정확한 음성 구간 검출이 요구된다. 이러한 음성 구간 검출 과정은 음성 다이얼링 시스템에서도 필수적이지만 자동차 소음이 존재하는 상황에서는 매우 어려워진다. 그림 2는 100 km/h의 속도로 고속도로를 주행하는 자동차의 햇빛 가리개에 설치한 마이크로 녹음한 음성 신호이다. 그림에서 알수 있듯이 음성의 시작점과 끝점을 정확히 검출하는 것은 어려워서 소음이 존재하지 않는 상황에서 사용되는 간단한 방법으로는 정확한 음성 구간 검출이 거의 불가능하다. 따라서 자동차 소음 환경에서 사용될 수 있는 음성 인식 시스템의 개발을 위해서는 소음 환경에서 사용될 수 있는 음성 구간 검출 알고리즘의 개발이 필수적이다.

본 연구에서는 자동차 소음 환경에서 정확한 음성 구간을 검출하기 위하여 그림 3과 같은 방법을 제안

하였다.

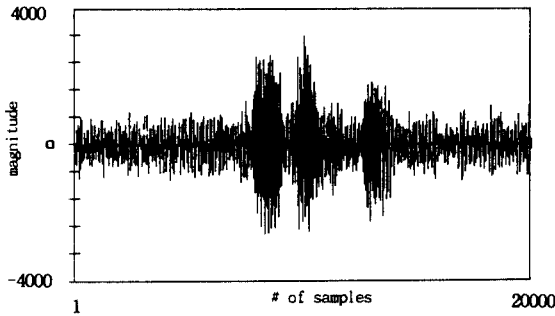


그림 2. 100 km/h로 주행중인 자동차에서 발음한 음성 '노상식'

Fig. 2. Time signal of the Korean name "no sang-sik", spoken in a car moving at 100km/h.

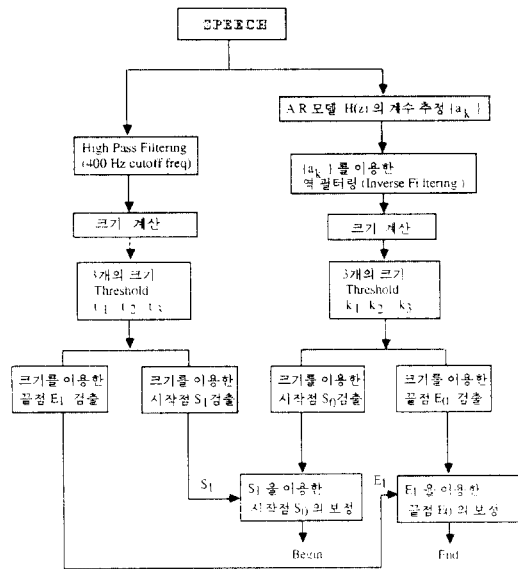


그림 3. 자동차 소음 환경에서의 음성 구간 검출 방법

Fig. 3. Endpoint detection method in the car noise environment.

음성 구간 검출 알고리즘은 크게 두 부분으로 나뉜다. 한 부분은 자동차 소음을 전극 필터로 모델링한 AR 계수를 이용하여 입력 신호를 역 필터링(inverse filtering)한 후, 신호의 크기를 이용하여 음성 구간 검출을 수행한다. 다른 부분은 자동차 소음의 대부분

이 저주파 영역에 집중되어 있으므로 400 Hz의 차단 주파수를 갖는 고역 통과 필터를 사용하여 소음을 제거한 신호의 크기를 이용하여 음성 구간 검출을 수행하고 역 필터링을 이용하여 구한 시작점과 끝점을 보정한다. 여기서 음성 구간 검출 방식은 3개의 임계값을 사용하는 전통적인 방법을 사용하였다.^[6] 이와 같은 방법은 소음의 스펙트럼을 모델링하여 잡음을 제거하므로 소음을 효과적으로 제거할 수 있고 잡음의 형태에 적응할 수 있는 장점이 있다. 그러나 잡음의 스펙트럼이 음성 신호와 비슷한 경우를 고려하여 역 필터링과 고역 통과 필터를 사용하는 방법을 병행한다.

3. 변형된 DTW

소음 환경에서 발생한 음성 구간 검출 오차는 조용한 환경에서 구한 기준 패턴과 소음 환경에서 발생한 시험 패턴을 비교하는 패턴 매칭 과정에 큰 영향을 미친다. 실제로 음성 구간 검출 과정에서 발생한 시작점과 끝점의 오차는 DTW 과정에서 어느정도 보상이 되지만, DTW 알고리즘은 두 패턴의 시작점과 끝점을 연결하는 최적 경로를 찾기 때문에 음성 구간 검출에서 발생하는 오차의 영향을 받게된다.

이러한 문제점을 해결하는 방법으로서, DTW 알고리즘을 시작점과 끝점의 오차를 고려하도록 변형하는 것이다. 이러한 방법을 그림 4에 나타내었다.^[7] 우선 시작점에서 발생하는 오차를 감소시키기 위하여 두 패턴 T와 R을 비교하기 이전에 처음 δs 프레임을 비교하여, 두 벡터간의 거리 $d(i(k), j(k))$ 가 최소가 되는 점을 두 패턴의 새로운 시작점으로 정한다. 식 (1)과 같은 이러한 과정은 시작점에 발생될수 있는 오차를 감소시킨다.

$$(i, j) = \underset{1 \leq i \leq \delta s, 1 \leq j \leq \delta s}{\operatorname{arg\,min}} d(i(k), j(k)) \quad (1)$$

끝점에서는 오차의 보상을 위하여 기준 패턴의 길이를 가변시킨다. 일반적으로 기준 패턴은 조용한 환경에서 생성되므로 소음이 존재하는 환경에서 발생하는 시험 패턴의 길이보다 길다. 따라서 DTW 알고리즘을 수행한 후에 끝점에서의 오차를 고려하여 기준 패턴의 길이를 δE 만큼 변화시켜보는 것이 필요하다. 두 패턴간의 최종 거리 $D(T, R)$ 는 누적 거리 $D_A(I, j)$ 를 이용하여 식 (2)과 같이 구한다.

$$D(T, R) = \min_{j=J-\delta E, \dots, J} D_A(I, j) \quad (2)$$

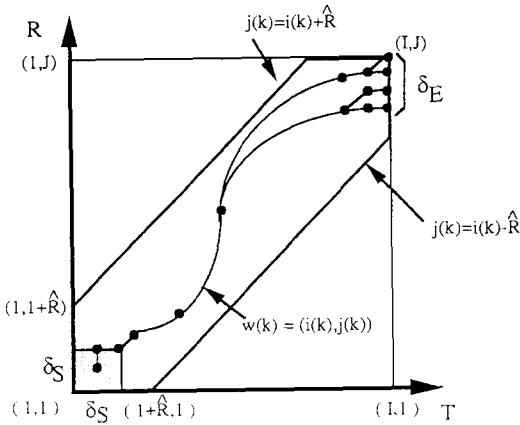


그림 4. 시작점과 끝점의 오차를 고려한 DTW 알고리즘

Fig. 4. Variants of DTW algorithm which take the starting and endpoint detection errors into consideration.

4. 잡음 제거와 LPC 분석

본 연구에서는 잡음 제거 방식으로 자기 상관 차감법 (autocorrelation subtraction method)을 사용하였다. [3] 소음 $n(k)$ 가 음성 신호 $s(k)$ 에 더해졌을 때, 소음 $n(k)$ 와 음성 신호 $s(k)$ 간에 상관 관계가 없다고 가정하면 개선된 자기 상관 함수 $r'_{ss}(\tau)$ 는 식 (3)과 같다.

$$r'_{ss}(\tau) = r_{ss}(\tau) - \alpha E[r_{nn}(\tau)], \quad 0 \leq \alpha \leq 1 \quad (3)$$

여기서 $E[r_{nn}(\tau)]$ 는 배경 잡음의 자기 상관 함수를 16 프레임(frame)동안 평균한 것이다. 따라서 소음의 특성이 시간적으로 변화하지 않는다고 가정하면 음성이 없는 묵음 구간에서 구한 소음의 자기상관 함수의 평균으로 음성에 섞인 소음 성분을 제거할 수 있다. 그러나 실제 상황에서는 잡음 성분의 지나친 제거로 인하여 개선된 자기 상관 함수 $r'_{ss}(\tau)$ 의 positive - definite 가 성립하지 않는 경우가 발생한다. [3] 이러한 문제점을 해결하기 위하여 특징 벡터를 구하는 선형 예측 분석(linear predictive analysis)과정에서 얻을 수 있는 잔차 에너지가 양수가 될때까지 α 값을 반으로 줄여 나갔다. 이러한 방법은 음성에 섞인 소음을 완전히 제거할 수는 없으나 계산이 간단하고 특징 벡터를 계산하는 중간 과정에서 이루어지므로 쉽게 구현될 수 있다.

5. 소음에 강한 거리 측정 방법

스펙트럼의 기울기에 기초를 둔 거리 측정은 시각에 의하여 음성간의 거리(perceived phonetic distance)를 측정하는 연구에서 Klatt에 의하여 제안되었고 [6]. 이러한 개념은 선형 예측 분석에 따른 계산 상의 잇점이 있는 전극 모델 스펙트럼에 직접 적용되었다. [2] 스펙트럼의 기울기(spectral slope distance measure)에 의한 거리 측정 방법은 식 (4)와 같이 표현된다.

$$dRPS = \sum_{k=1}^P [k(c_k - c'_k)]^2 \quad (4)$$

여기서 c_k, c'_k 는 커스트럼 계수이다. 이러한 거리 측정 방법의 특징은 소음의 영향을 비교적 적게 받는 포먼트의 변화에 민감하기 때문에 소음이 존재하는 상황에서 안정된 거리 측정 방법이다.

III. 실험 및 결과

위에서 기술한 음성 인식 시스템은 모토로라 DSP56001(24 비트 범용 신호처리 프로세서)를 사용하여 구현되었으며 자동차의 햇빛 가리개에 설치한 마이크(AT831b)로 녹음한 음성 데이터를 사용하여 화자 종속 실험을 수행하여 성능을 평가하였다.

1. 데이터 베이스

음성 인식에 사용되는 데이터 베이스는 각각 50개의 인명(name)으로 이루어진 두개의 집단(GROUP-1, GROUP-2)으로 구성된다. 시험 패턴은 100km/h로 고속도로를 주행중인 자동차에서 3명의 20대 남성 화자가 각 단어를 2회씩(50개 * 2집단 * 2회 * 3명=600개) 발음한 음성으로 구성되고, 기준 패턴은 정지한 자동차에서 엔진을 끈 상태와 켜진 상태에서 각각 1회씩 발음한 음성을 사용하였다. 시험 패턴의 소음 레벨은 평균 63 dBA 였다.

2. 음성 인식 시스템의 구성

실험에 사용된 음성 인식 시스템의 블록도는 그림 1과 같다. 음성 신호는 8 kHz, 16 비트로 표본화된 후, 전처리 과정에서 250Hz의 차단 주파수를 갖는 고역 통과 필터를 거친다. 이렇게 샘플링된 신호는 음성 구간과 묵음 구간을 구별하기 위하여 음성 구간 검출을 수행하고 특징 벡터를 구한다. 이때 소음의 영향을 고려하여 자동차 소음의 영향을 고려한 음성 구간 검출 알고리즘을 사용하였으며, 자기 상관 차감법을 사용하여 잡음을 제거하였다. 배경 잡음 레벨

추정 과정에서는 약 200ms 정도의 시간 동안에 배경 잡음의 크기, 스펙트럼의 형태와 그것의 자기상관 함수를 추정하게 된다. 이러한 파라미터를 바탕으로 소음의 영향을 최소화하여 끝점 검출을 수행한다. 검출된 음성 신호는 32ms(256샘플)의 크기를 갖는 해밍(hamming) 윈도우를 사용하여 16ms씩 이동하면서 차수 P=14인 선형 예측 계수를 구하는 LPC 분석 과정을 거친다. 이때 자기 상관 차감법을 수행하여 잡음의 영향을 제거한 후에 선형 예측 계수를 구한다. 이러한 LPC 계수로 부터 인식 과정에 사용될 LPC 커스트럼 계수를 선형 예측 계수와 동일 차수까지 순환식을 이용하여 구한다. 결정 법칙(decision rule)은 비교된 결과를 각 단어당 기준 패턴수를 고려하여 최종 인식을 결정하는 단계로서 NN(Nearest Neighbor) 법칙을 사용하였다.

3. 음성 인식 시스템 구현

본 연구에서 구현한 음성 인식 시스템은 패턴 매칭 구조를 기본으로 하는 화자 종속 단독음 인식 시스템으로서, 범용 신호 처리 칩인 DSP56001을 사용하는 PC 보드(board)를 이용하여 DSP56001의 어셈블러로 구현하였다. 음성 인식 시스템의 구성은 음성 인식을 수행하는 DSP 시스템 보드, 사용자 인터페이스와 DSP 시스템을 제어하는 IBM PC 로 구성된다.

음성 인식 소프트웨어의 구성은 음성 인식 시스템 학습, 인식 그리고 인식 시스템 제어의 세부분으로 나뉘어 진다. 그림 5와 6에서는 학습 모드(training mode)와 인식 모드(recognition mode) 상태에서의 소프트웨어 구성을 보여준다.

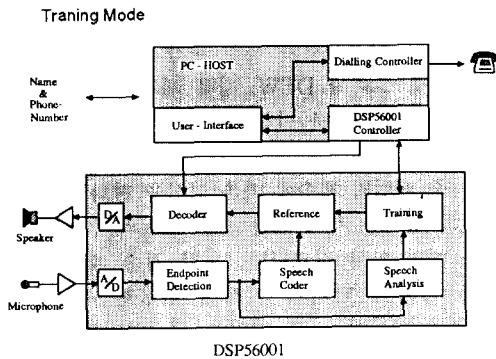


그림 5. 학습 모드에서의 시스템 소프트웨어 구성도

Fig. 5. Block diagram of system software configuration in the training mode.

그림 5는 학습 모드에서의 인식 시스템 상황을 나타낸다. 마이크를 통하여 들어온 입력 음성은 음성 구간 검출, 잡음 제거와 음성 분석 과정을 거쳐서 특징 벡터의 형태로 메모리에 기억된다. 이와 동시에 입력된 음성 신호도 저장되어 인식 결과를 확인하는데 사용된다. 이때 PC 호스트(host)에서는 인식 시스템에 저장되는 기준 패턴의 상태를 파악하고 사용자와의 인터페이스를 담당하며 각각의 기준 패턴에 해당하는 전화번호를 키보드를 통하여 받아들인다.

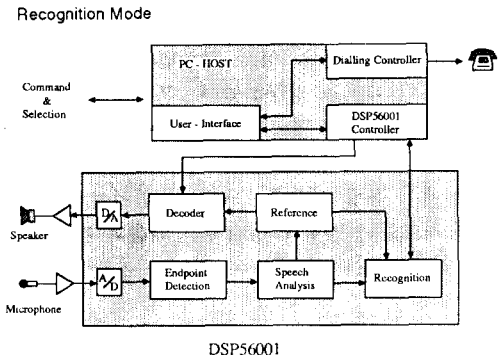


그림 6. 인식 모드에서의 시스템 소프트웨어 구성도

Fig. 6. Block diagram of system software configuration in the recognition mode.

그림 6은 인식 모드에서의 인식 시스템 상황을 나타낸다. 입력 음성 신호는 학습 모드에서와 마찬가지로 음성 구간 검출, 잡음 제거와 음성 분석을 거쳐서 특징 벡터로 변환된다. 변환된 특징 벡터는 기준 패턴과 비교되어 인식 결과를 PC 호스트로 전달한다. 인식 결과는 3개까지의 후보를 선정하여, 인식 결과가 틀린 경우에 다음 후보를 선정할 수 있도록 한다.

시스템 제어 프로그램은 인식 시스템에게 음성을 인식할 수 있는 명령을 내리고, 수행한 결과를 받아들이고 사용자와의 인터페이스와 인식 시스템의 상태를 제어한다. 시스템 제어 프로그램이 인식 시스템에게 전달하는 명령은 표준 패턴의 저장, 인식, 인식 시스템 초기화, 시스템의 상태 검사, 명령 취소, 표준 패턴을 다운 로드(down load), 표준 패턴을 업 로드(up load), 표준 패턴의 삭제등이다. 예로서 인식의 경우, 제어 시스템이 인식 시스템에게 인식 수행 명령을 전달하면 인식 시스템은 인식을 수행하여 인식 결과를 3개의 후보까지 제어 시스템에게 전달된다.

이때 제어 프로그램은 사용자가 인식한 결과를 확인할 수 있도록 첫번째 후보에 해당되는 단어의 음성 신호를 출력하도록 인식 시스템에게 명령하고 음성 신호를 인식 시스템에게 전달한다. 사용자는 인식한 단어가 정확할 경우에는 음성 인식 시스템에게 다이얼링을 명령하고 틀린 경우에는 다음 후보의 단어를 출력하도록 명령하거나 취소할 수 있도록 한다. 이렇게 확인된 음성의 전화번호는 다이얼링 콘트롤러(dialing controller)를 통하여 다이얼링하게 된다.

4. 음성 인식 시스템의 형태에 따른 성능 평가

본 연구에서는 자동차 소음 환경에서 음성 인식을 수행하기 위하여 음성 인식 시스템의 전반에 걸쳐서 소음에 대한 대책을 강구하였다. 본 절에서는 이러한 방법들의 유효성을 평가하기 위하여 각각의 방법을 사용하는 음성 인식 시스템의 성능을 비교하였다. 소음 환경에서 음성 인식을 수행하기 위하여 본 연구에서 사용하고있는 방법을 정리하면 다음과 같다.

- (1) 역 필터링과 고역 통과 필터 사용한 음성 구간 검출.
- (2) 자기 상관 차감법을 사용한 잡음 제거.
- (3) 소음에 영향을 적게 받는 거리 측정 방법.
- (4) 음성 구간 검출 오차를 고려한 변형된 DTW.

표 1. 여러가지 형태의 음성 인식 시스템과 파라메타

Table 1. Several types of speech recognizer and their parameters.

방법 \ 형태	CASE 1	CASE 2	CASE 3	CASE 4	CASE 5
역 필터링과 고역 통과 필터	○	○	○	x	○
자기 상관 차감법	○	x	○	○	○
음성 구간 검출 오차를 고려한 DTW	○	○	x	○	○
스펙트럼 기술기에 의한 거리 측정	○	○	○	○	x

위와 같은 방법들의 유효성을 평가하고 가장 우수한 성능을 나타내는 방법을 선정하기 위하여 표 1와 같은 5가지 형태의 음성 인식 시스템을 구현하였다. CASE 1은 본 연구에서 소음에 대한 대책으로 제시한 모든 방법을 사용하는 경우이다. CASE 2는 잡음에 대한 4가지 대책중에서 입력 신호에 포함된 잡음을 제거하는 자기 상관 차감법을 사용하지 않는 시스템이다. CASE 3의 경우에는 음성 구간 검출 오차를 고려한 DTW 를 사용하지 않고 시작점과 끝점이 고정된 일반적인 DTW 방식을 사용한 경우이다.

CASE 4는 자동차 소음 환경에서 음성 구간의 정확한 검출을 위하여 제안된 역 필터링을 수행하지 않고 고역 통과 필터만을 사용하여 음성 구간 검출을 수행하는 경우이다. CASE 5는 거리 측정 방법에 따른 음성 음성 인식 시스템의 성능을 평가하기 위한 방법으로 일반적으로 사용되고 있는 커스트럼 거리 측정 방법(cepstral distance measure)^[9]을 사용한 시스템이다. 즉, CASE 2,3,4,5의 결과는 CASE 1의 결과와 비교하여 소음에 대한 대책으로 강구된 각 방법의 유효성을 알 수 있다.

표 2. 음성 인식 시스템의 형태에 따른 인식율
Table 2. recognition rates of several speech recognizers according to their configuration.

대상\형태	CASE 1	CASE 2	CASE 3	CASE 4	CASE 5
GROUP-1	96.00	96.00	95.00	93.50	93.00
GROUP-2	98.50	97.50	97.50	95.00	95.00
평균	97.25	96.75	96.25	94.25	94.00

표 2는 5가지 경우에 대한 인식 실험 결과를 보여 준다. 인식 실험은 2명을 대상으로 각 단어당 하나의 기준 패턴을 사용하였다. 그림 5에서 가장 인식율이 높은 방법은 CASE 1인 경우이다. 즉 본 연구에서 제시한 소음에 대한 4가지 방법을 모두 사용한 경우가 97.25%로 가장 높은 인식율을 나타내었다. 또한 CASE 2,3,4,5를 CASE 1과 비교하여 볼때, 인식율의 차이는 CASE 5,4,3 2의 순서로 차이가 크다. 이것은 4가지의 소음에 대한 대책중에서 거리 측정 방법이 가장 큰 역할을 한다는 것을 의미한다. 그 다음으로는 소음 환경에서의 음성 구간 방법, 음성 구간 검출 오차를 고려한 DTW, 잡음 제거의 순서로 음성 인식 시스템에 영향을 미친다.

5. 기준 패턴의 개수에 따른 인식율 변화

표 2의 결과는 각 단어당 기준 패턴의 수를 한개로 하였을 때의 인식 결과이다. 또한 기준 패턴의 선택 과정도 각각의 단어를 여러번 발음한 후 집단화(clustering)하여 기준 패턴을 생성하지도 않았다. 따라서 각 단어당 하나의 기준 패턴을 생성하는 것보다는 두개 이상을 생성하여 음성의 변화에 보다 잘 대응할 수 있도록 하는 것이 필요하다.

본 절에서는 각 단어당 기준 패턴의 수가 한개일 경우와 두개인 경우에 음성 인식 시스템의 성능을 비교하였다. 각 단어당 한개의 기준 패턴을 갖는 경우,

자동차가 정지한 상태에서 엔진을 끄고 발음한 음성 과 엔진을 켜고 발음한 음성을 각각 기준 패턴으로 사용하였으며, 각 단어당 두개의 기준 패턴을 갖는 경우에는 두 상태에서 발음한 음성을 한 단어에 대한 기준 패턴으로 저장하여 인식을 수행하였다.

표 3. 단어당 기준 패턴의 갯수에 따른 인식율 변화

Table 3. recognition rate of speech recognizer according to the number of reference pattern for each word.

기준 패턴 대상	ref 1(off)		ref 1(on)		ref 2	
	1st	2nd	1st	2nd	1st	2nd
GROUP-1	96.00	98.67	97.00	99.00	99.67	99.67
GROUP-2	94.33	98.00	95.33	99.33	97.33	98.67
평균	95.17	98.67	96.17	99.17	98.50	99.17

표 3은 3명을 대상으로 수행한 기준 패턴의 갯수에 따른 인식율 변화로서, 단어당 한개의 기준 패턴을 사용했을 때(ref 1(on), ref 1(off))와 두개를 사용했을 때(ref 2)의 인식율을 나타낸다. 단어당 한개의 기준 패턴을 사용한 경우, 엔진을 켜 상태에서 생성한 기준 패턴을 사용한 경우(ref 1(on))의 인식율은 96.16%로서 엔진을 끈 상태(ref 1(off))에서 생성한 경우보다 1% 정도 높은 인식율을 나타낸다.

표 4. 음성 인식 시스템의 특성

Table 4. Main features of the speech recognizer.

특 성	내 역
인식 형태	화자 종속 단독음
기준 패턴당 단어수	1개 이상 (사용자 정의 가능)
인식율	98.5% - 이음 50개, 남자 3명, 단어당 기준 패턴 2개 - 인식 상황 : 100 km/h로 주행중인 자동차 환경 - 소음 레벨 : 35 - 45 dBA (학음 모드) 63 dBA (인식 모드)
단어수	약 100초(이음 100개 정도)
응답 시간	1 초 이내 (100 단어 경우) (단어의 발음이 끝나는 순간부터 호스트로 응답이 오는데 걸리는 시간.)
유효 발음 길이	0.2 - 3.1초
화자의 입과 마이크간의 거리	40 cm

이러한 결과는 고속 주행중에 발음한 각 화자의 시험 패턴이 엔진을 끈 상태보다 켜 상태에서 발음한 것에 더욱 유사하다는 것을 나타낸다. 또한 각 단어당 두

개의 단어를 사용한 ref 2의 경우에는 각 단어당 한 개의 기준 패턴을 사용한 경우보다 더 높은 98.5%의 인식율을 나타내었다. 이러한 결과는 같은 시험 패턴에 대하여 ref 1(off)과 ref 1(on)의 경우에 각각 발생하던 오차가 각 단어당 두개의 기준 패턴을 사용함으로써 상호 보완적으로 작용하여 전체적인 인식율을 높인 것이다. 따라서 2개 이상의 기준 패턴을 사용하여 음성의 변화에 대응할 수 있는 것이다. 음성 인식 과정에서 발생하는 대부분의 오차는 단어의 형태가 비슷한 경우와 배경 잡음의 상태가 급격히 변화하는 경우에 발생하였다. 오차가 많이 발생한 단어는 '이재영' 과 '이대영', '김영훈' 과 '임용훈', '박영철' 과 '박용철' 등이다.

IV. 결론

본 연구에서는 자동차 소음 환경에서 사용될 수 있는 화자 종속 음성 인식 시스템의 구현에 관하여 기술하였다. 이러한 목적을 위하여 자동차 소음 환경에서 음성 인식 시스템의 성능을 개선할 수 있는 방법들을 제시하였고, 제안된 음성 인식 시스템은 범용 신호 처리 프로세서인 모토로라사의 DSP56001을 사용하여 구현하였다. 음성 인식 시스템의 성능 평가는 자동차 주행중에 발음한 음성 데이터를 사용하였다.

본 연구에서는 자동차 소음 환경에서 음성 인식 시스템의 성능을 개선하기 위하여 다음과 같은 방법을 사용하였다. 첫째, 자동차 소음 환경에서 음성 구간 검출을 수행하기 위하여 역 필터링과 고역 통과 필터를 통과한 신호의 크기를 파라미터로 하는 음성 구간 검출을 수행하였다. 또한 패턴 매칭 과정에서도 변형된 DTW 알고리즘을 사용하여 음성 구간 검출 오차를 보상하였다. 둘째, 음성에 포함된 자동차 소음을 제거하기 위하여 선형 예측 과정에서 얻을 수 있는 잔차 신호의 에너지가 양수가 되도록 하는 제한 조건을 갖는 자기 상관 차감법을 사용하였고, 소음에 강한 거리 측정 방법을 사용하여 소음의 영향을 최소화 하였다.

본 연구에서 구현한 음성 인식 시스템의 성능과 소음에 대한 영향을 평가하기 위하여 100 km/h 로 주행중인 자동차에서 녹음한 음성을 사용하여 인식 실험을 수행하였다. 구현한 음성 인식 시스템의 특성은 표 4에 나타내었다. 음성 인식 실험을 수행한 결과, 소음에 대한 영향을 가장 적게 받는 경우는 본 연구에서 제한한 소음에 대한 대책을 모두 사용하는 경우였다. 이때 각 단어당 한개의 기준 패턴을 사용한 경우의 인식율은 95.67%였고 각 단어당 2개를 사용한

경우의 인식율은 98.5% 였다.

参 考 文 献

- [1] H.W. Ruehl, S. Dobler, J. Weith, P. Meyer, "Speech Recognition in the Noisy Car Environment," *Speech Communication*, vol. 10, no. 1, pp. 11-22, Feb. 1991
- [2] B.A. Hanson, H. Wakita, "Spectral Slope Distance Measure with Linear Prediction Analysis for Word Recognition in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, no. 7, pp. 968-973, July 1987
- [3] S. Kay, "Noise Compensation for Autoregressive Spectral Estimation," *IEEE Trans., Acoust., Speech and Signal Processing*, vol. ASSP-28, pp. 292-303, June 1980
- [4] C. Mokbel, G. Chollet, "Word Recognition in the Car," ICASSP-91, pp.925-928
- [5] DSP56000/DSP56001 Digital Signal Processor User Manual, MOTOROLA Inc., 1990
- [6] L.F. Lamel, L. R. Rabiner, "An Improvement Endpoint Detector for Isolated Word Recognition," *IEEE Trans., Acoust., Speech and Signal Processing*, vol. ASSP-29, no.4, pp. 777-785, Aug. 1981
- [7] L. R. Rabiner, A. E. Rosenberg, S. E. Levinson, "Considerations in Dynamic Time Warping Algorithms for Discrete Word Recognition", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 575-582, Dec. 1978
- [8] D.H. Klatt, "Prediction of Perceived Phonetic Distance from Critical Band Spectra : A first step", Proc. ICASSP-82, pp. 1278-1281, May 1982
- [9] A. H. Gray, Jr, J. D. Markel, "Distance Measures for Speech Processing", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, no. 5, pp. 380-391, Oct. 1976
- [10] 김락용, 김원구, 임용훈, 차일환, 윤대회, "백색 및 유색 잡음 환경에서의 단독음 인식," 전자공학회논문지, 제 28권 B편, 제 6호, pp. 454- 461, 1991년 6월

著 者 紹 介



金元九 (正會員)
1964年 4月 3日生. 1987年 2月 연세대학교 전자공학과 졸업, 1989年 8月 연세대학교 전자공학과 석사 학위 취득, 1989年 9月 ~ 현재 연세대학교 전자공학과 박사과정 재학 중 주관심 분야는 디지털 신호처리 및 음성 신호처리 등임.

尹大熙 (正會員) 第 28卷 第 10卷 3號 參照.
현재 연세대학교 전자공학과 교수

車日煥 (正會員) 第 28卷 第 10卷 3號 參照.
현재 연세대학교 전자공학과 교수.