

# 자동차 환경에서의 단독 숫자음 및 명령어 인식

## Isolated Digit and Command Recognition in Car Environment

양 태 영\*, 신 원 호\*, 김 지 성\*\*, 안 동 순\*\*\*, 이 충 용\*, 윤 대 회\*, 차 일 환\*  
 (Tae Young Yang\*, Won Ho Shin\*, Ji Sung Kim\*, Dong Soon Ahn\*\*,  
 Chung Yong Lee\*, Dae Hee Youn\*, and Il Whan Cha\*)

\*본 논문은 1997년 현대전자산업(주)의 지원으로 연구되었습니다.

### 요 약

본 논문에서는 DHMM(Discrete Hidden Markov Model) 기반의 음성 인식 시스템에서 소음에 강인한 인식 성능을 얻기 위하여, 관찰 확률 스무딩(observation probability smoothing) 방법을 제안하고, 자동차 소음하에서의 음성 인식에 적합한 소음처리 기법을 실험을 통해 제시한다. 제안된 관찰 확률 스무딩 방법은 입력되는 음성의 특징벡터가 소음에 오염되어 양자화(vector quantization) 과정에서 적절한 코드워드(codeword)가 아닌 다른 코드워드로 양자화됨으로써 발생하는 인식 성능 저하를 막기 위하여, 각각의 코드워드와 거리가 가까운 코드워드들의 관찰 확률값을 높여주는 방법이다. 이 밖에 자동차 소음에 대한 대처 방안으로 특징 벡터의 거리 측정시의 리프터(lifter) 사용, 고역 통과 필터(high pass filter) 사용, 스펙트럴 차감법(spectral subtraction) 사용 등의 성능을 평가한다. 인식 실험은 자동차 정지 중과 주행 중의 두 가지 상황에서 녹음된 한국어 단독 숫자음과 명령어 14단어에 대해 수행하였으며, 정지 중 97.4%와 주행 중 59.1%의 인식률로부터, 제안된 관찰 확률 스무딩 방법과 리프터, 고역 통과 필터, 스펙트럴 차감법의 소음 처리 기법을 추가한 결과, 정지 중 98.3%와 주행 중 88.6%의 인식률을 얻을 수 있었다.

### ABSTRACT

This paper proposes an observation probability smoothing technique for the robustness of a discrete hidden Markov (DHMM) model based speech recognizer. Also, an appropriate noise robust processing in car environment is suggested from experimental results. The noisy speech is often mislabeled during the vector quantization process. To reduce the effects of such mislabelings, the proposed technique increases the observation probability of similar codewords. For the noise robust processing in car environment, the liftering on the distance measure of feature vectors, the high pass filtering, and the spectral subtraction methods are examined. Recognition experiments on the 14-isolated words consists of the Korean digits and command words were performed. The database was recorded in a stopping car and a running car environments. The recognition rates of the baseline recognizer were 97.4% in a stopping situation and 59.1% in a running situation. Using the proposed observation probability smoothing technique, the liftering, the high pass filtering, and the spectral subtraction, the recognition rates were enhanced to 98.3% in a stopping situation and to 88.6% in a running situation.

### I. 서 론

DHMM(Discrete Hidden Markov Model)을 기반으로 한 음성 인식 시스템에서 특징벡터(feature vector)의 양자화(vector quantization) 과정은 인식 성능에 큰 영향을 미친다. DHMM의 학습 과정에서 관찰 확률(observation probability)

은 학습 데이터(training data) 특징벡터의 발생 정도에 따라 주어지게 되므로, 인식 과정에서 입력 특징벡터 중에 관찰 확률이 매우 낮은 코드워드(codeword)로 양자화되는 특징 벡터가 있을 경우, 전체 확률에도 큰 영향을 주게 되어 인식 성능을 저하시킬 수 있다. 소음이 존재하는 환경에서의 음성 인식에서도 같은 문제가 발생한다. 소음에 오염된 입력 음성의 특징벡터는 학습 과정에서 많이 관찰되던 주요 코드워드가 아닌 관찰 확률이 낮은 코드워드로 양자화될 확률이 크기 때문에 이로 인한 많은 오류를 일으킬 수 있다. 따라서 이러한 문제를 해결할 수 있는 관찰 확률 스

\* 연세대학교 전자공학과

\*\* (주) 로커스(Locus)

\*\*\* 목포대학교 컴퓨터공학과

접수일자: 1998년 9월 3일

무딩(observation probability smoothing) 기법이 요구된다.

기존의 관찰 확률 스무딩은 주로 충분하지 못한 학습 데이터로 인해 코드북(codebook) 각 코드워드의 정확한 관찰 확률을 구하기가 어렵다는 가정 아래, 예상하지 못한 코드워드, 특히 학습 과정에서 한 번도 관찰되지 못하여 관찰 확률값이 "0"인 코드워드가 선택될 경우에 대비하기 위한 방법으로서 연구되었다. 가장 대표적인 방법으로 관찰 확률값이 "0"이면 이 보다 큰 작은 값으로 대체시키는 플로어 스무딩(floor smoothing) 방법[1]이 제안되었으며, 인식 시스템의 학습 과정에서 각 코드워드의 거리를 고려하여 관찰 확률값을 설정하는 거리 스무딩(distance smoothing) 방법[2][3]도 사용되었다. 또한, 동일한 모델을 학습시키기 위한 데이터들에서 특정 코드워드가 다른 코드워드로 대체되어 나타나는 빈도를 조사하여 관찰 확률을 스무딩하는 동반발생 스무딩(co-occurrence smoothing) 방법[4][5]도 있다.

본 논문에서는 소음으로 인해 입력 음성의 특징벡터가 알맞은 코드워드로 정확히 양자화되지 못하는 경우에 대비해 코드북(codebook)에서 서로 근접한 거리에 위치한 코드워드들의 관찰 확률 값을 높여 주는 관찰 확률 스무딩 기법을 제안한다. 제안된 스무딩 기법은 소음에 오염된 특징벡터가 양자화되는 코드워드가 소음이 없을 경우 양자화될 코드워드와 매우 먼 거리에 있을 경우보다는 근접한 거리에 있을 확률이 크다는 가정에서 출발하며, 기존 스무딩 방법들의 목적과는 달리 소음에 대처하기 위한 스무딩 방법이므로 기존의 스무딩 방법에 비해 스무딩에 사용되는 확률값이 크고, 코드워드들 사이의 가까운 순서에 비례하여 스무딩 확률값의 크기가 결정된다.

이와 함께 본 논문의 인식 실험 환경인 자동차 환경에서의 인식 성능 향상을 위하여 자동차 주행 소음에 대처하기 위한 방안으로 특징 벡터의 거리 측정시의 리프터(lifter)[6]-[10] 사용, 고역 통과 필터링(high pass filtering), 스펙트럴 차감법(spectral subtraction)[11][12] 등의 성능을 평가하고, 적절한 소음 대처 방법을 제시한다.

본 논문의 구성은 다음과 같다. 제 II장에서는 제안한 관찰 확률 스무딩에 대해 설명하고, 제 III장에서 인식 실험을 위한 전처리 및 특징벡터 추출 과정과 인식 시스템의 구성을 설명하며, 제 IV장에 인식 실험에 사용된 데이터 베이스를 소개한다. 인식 실험의 과정과 그 결과는 제 V장에서 제시하며, 제 VI장에 본 논문의 결론을 맺는다.

II. 관찰 확률 스무딩

소음이 존재하는 환경에서의 음성 신호는 조용한 환경에서의 음성 신호에 비해 인식률이 낮다. DHMM을 기반으로 한 음성 인식 시스템의 경우, 이것은 소음에 오염된 음성 신호의 양자화 결과가 깨끗한 음성 신호의 양자화 결과와 다르다는 점이 주된 원인이라 할 수 있다. 소음에 오염된 음성 신호가 알맞은 코드워드로 양자화되지 않는 때에는, 소음의 정도에 따라 차이가 있지만 그와 유사한 코드워드로 양자화되는 경우도 많다. 본 논문에서는 이러

한 상황에서의 인식 성능 저하를 막기 위하여, 코드북의 각 코드워드를 기준으로 그와 인접한 거리에 있는 코드워드들을 검색하고, 검색된 코드워드들의 관찰 확률을 높여주는 스무딩 방법을 제안한다.

전체  $L$ 개의 코드워드로 구성된 코드북을 갖는 DHMM의 각 모델(model)이  $N$ 개의 상태(state)로 이루어져 있고,  $j$ 번째 상태에서  $l$ 번째 코드워드가 관찰될 확률을  $b_j(l)$ 이라 할 때, 제안된 관찰 확률 스무딩 과정은 다음의 3단계로 이루어진다.

과정-1>  $l$ 번째 코드워드와 가장 가까운 거리에 있는  $K$ 개의 코드워드를 찾는다.

과정-2> 과정-1에서 검색된  $k$ 번째로 가까운 코드워드의 순번(index)을  $i(l, k)$ 라 하면, 다음과 같은 비교 과정을 통해 관찰 확률을 스무딩한다.

$$b_j(i(l, k)) = \max[b_j(i(l, k)), (\alpha - (k-1)\beta) b_j(l)], \quad 1 \leq j \leq N, 1 \leq l \leq L, 1 \leq k \leq K. \quad (1)$$

여기서  $\alpha$ 와  $\beta$ 는 상수이다.

과정-3> 과정-2에서 변화된 관찰 확률들을

$$\sum_{i=1}^K b_j(l) = 1, \quad 1 \leq j \leq N. \quad (2)$$

의 확률 조건이 만족하도록 다음과 같이 정규화시켜 준다.

$$\delta_j(l) = \frac{b_j(l)}{\sum_{i=1}^K b_j(l)}, \quad 1 \leq j \leq N, 1 \leq l \leq L. \quad (3)$$

제안된 관찰 확률 스무딩은 하나의 코드워드에 대해 가장 가까운  $K$ 개의 코드워드의 관찰 확률에 삼각형 모양으로 최소값을 주는 것이다. 상수  $\alpha$ 는 스무딩을 위한 삼각형의 꼭지점 즉, 스무딩의 초기값을 결정하는 상수이며,  $\beta$ 는 스무딩을 위한 삼각형의 기울기를 결정하는 상수이다. 제안된 관찰 확률 스무딩 과정의 개념을 그림 1에 나타내었다.

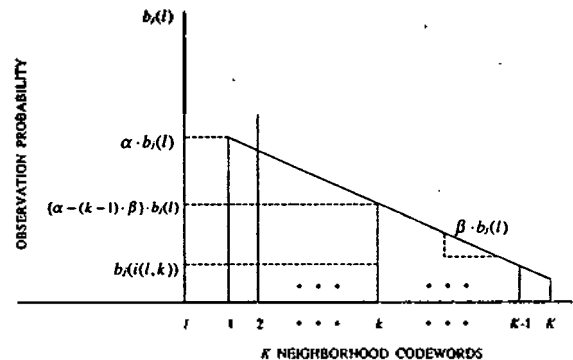


그림 1. 관찰 확률 스무딩 과정  
Fig. 1. Observation probability smoothing process.

### III. 인식 시스템의 기본 구성

#### 3.1 전처리 및 특징벡터 추출 과정

음성 신호는 8 kHz 16 bit로 샘플링되어  $1-0.95z^{-1}$ 의 전달 함수를 갖는 프리엠퍼시스(pre-emphasis) 필터를 거친 후, 20 ms(160 샘플)의 길이를 갖는 해밍윈도우(Hamming window)를 사용하여 10 ms씩 이동하면서 분석한다. 각 프레임(frame)마다 12차 MFCC(Mel-Frequency Cepstral Coefficient)와 에너지를 구하고 현재 프레임을 중심으로 2프레임 간격으로 델타 MFCC(delta MFCC), 델타 에너지(delta energy), 델타 델타 에너지(delta-delta energy)를 구하여 음성 인식 시스템의 특징벡터(feature)로 사용하였다.

#### 3.2 HMM 구성

자동차 환경에서의 단독 숫자음 및 명령어 인식을 위하여 본 논문에서는 DHMM을 기반으로 한 화자 독립 단독음 인식 시스템을 구현하였다. 단어 단위 인식 모델을 구성하였으며, 각 인식 대상 단어는 하나의 모델을 갖는다. 각 모델은 단어에 포함되는 음소의 수에 따라 7-12개의 상태를 갖도록 하였다. 주행중인 차량과 같은 소음이 심한 환경에서의 음성 인식의 경우, 음성 구간과 소음 구간을 구별하는 것은 매우 어려우며, 음성 구간을 검출하여 인식하는 것은 더 많은 오인식을 유발할 수 있다. 따라서 본 논문에서 사용된 단어 모델은 연속음 인식을 응용하여 각 모델의 전후에 소음 모델이 결합된 모델을 사용하였으며, 그 결과 모든 단어 모델의 첫 번째 상태와 마지막 상태는 모두 동일한 상태로 구성되었다. 이 소음 모델은 연속음 인식에서와 같이 사용되거나 생략되는 것이 자유롭기 때문에 음성 신호 전후에 소음이 많이 있는 상황이나 소음이 없이 음성 신호만이 입력된 상황에서도 인식 성능에 크게 영향을 주지 않는다. 사용된 모델의 한 예를 그림 2에 보이고 있다.

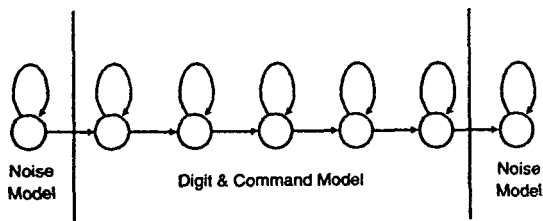


그림 2. 소음 모델이 전후에 결합된 단어 모델의 예  
Fig. 2. An example of a word model with noise models.

DHMM에 사용되는 코드북은 위의 특징벡터 추출 과정으로부터 얻은 학습 데이터의 특징벡터들을 LBG 알고리즘을 사용하여, MFCC와 델타 MFCC의 경우 256개의 코드워드, 델타 에너지와 델타-델타 에너지를 붙여 사용한 특징벡터의 경우 64개의 코드워드를 갖도록 구성하였다. DHMM의 학습과정에서 기본적으로 줄로어 스무딩 방법

[1]에 의한 관찰 확률 스무딩을 수행하였으며, MFCC와 델타 MFCC의 특징벡터에 대한 스무딩 값은  $1.0 \times 10^{-4}$ 로 하였고, 델타 에너지와 델타-델타 에너지를 붙여 사용한 특징벡터에 대한 스무딩 값은  $5.0 \times 10^{-4}$ 로 하였다.

### IV. 데이터 베이스(Data Base)

본 논문의 인식 대상은 0(영)에서부터 9(구)까지의 10개의 숫자음에 0을 “공”으로 발음한 경우를 합한 11개의 한국어 단독 숫자음과 전화를 걸 때 필요한 “다시”, “취소”, “걸어”의 3가지 명령어로 구성된다. 이것을 표 1에 나타내었다.

표 1. 인식 대상 단어  
Table 1. Recognition word set.

숫자음 (11단어)	영, 일, 이, 삼, 사, 오, 육, 칠, 팔, 구, 공
명령어 (3단어)	다시, 취소, 걸어
총 14 단어	

인식 시스템의 학습을 위해서 남성 20명과 여성 20명이 시동을 끄고 정지해 있는 차량에서 각 인식 대상 어를 5번씩 발음한 총 2,800 단어를 사용하였고, 인식 성능의 평가를 위해서 남성 5명과 여성 5명이 시동을 켜고 정지해 있는 차량에서 각 단어를 3번씩 발음한 420 단어와 100 km/h로 주행 중인 차량에서 3번씩 발음한 420 단어를 사용하였다. 음성 신호 녹음에 사용된 마이크는 Audio Technica사의 AT831b이며, 차량 조수석의 햇빛 가리개에 부착하여 녹음하였다.

### V. 인식 실험 및 결과

본 장에서는 자동차 환경에서의 단독 숫자음 및 명령어 인식을 위한 기본 인식 시스템의 성능과 자동차 소음에 대처하기 위한 소음 처리 기법들을 기본 인식 시스템에 추가해 가는 과정들에서의 성능 향상을 다룬다. 기본 인식 시스템의 구성은 제 III장에서 설명된 바와 같고, 자동차 소음 처리 기법으로 리프터링(liftering), 고역 통과 필터(high pass filter: HPF), 제안된 관찰 확률 스무딩(observation probability smoothing: OPS) 방법, 스펙트럴 차감법(spectral subtraction: SS)을 사용하였다.

#### 5.1 기본 인식 시스템의 인식 결과

기본 인식 시스템의 자동차 정지 중 및 주행 중의 인식 성능은 다음 표 2와 같다. 표 2에서 “STOP”은 정지 중에 녹음된 시험 데이터에 대한 인식률을 나타내며, “RUN”은 주행 중에 녹음된 시험 데이터에 대한 인식률을 나타낸다. 소음이 없는 정지 상태에서의 인식률은 우수하나, 자동차 주행 소음이 있는 상황에서는 인식률이 크게 저하됨을 알 수 있다.

표 2. 기본 인식 시스템의 인식률 [%]  
Table 2. The recognition rates of a baseline system.

	STOP	RUN
RECOGNITION RATE [%]	97.4	59.1

5.2 소음 처리 1 : Liftering

자동차 주행 소음에 대한 대비책 중 하나로 잡음에 강한 특징벡터의 거리측정 방법을 기본 인식 시스템에 추가하였다. 기본 인식 시스템의 거리측정 방법인 유클리디안(Euclidean: EUC) 거리측정 방법과 잡음 환경에서 좋은 성능을 보이는 것으로 알려져 있는 가중 캡스트럴 거리측정 방법인 RPS(Root Power Sum)[6][7], SLL(Smoothed Linear Lifter)[8], BPL(Band Pass Lifter)[9], GEL(General Exponential Lifter)[10]에 대한 인식 성능을 비교하여 그림 3에 나타내었다.

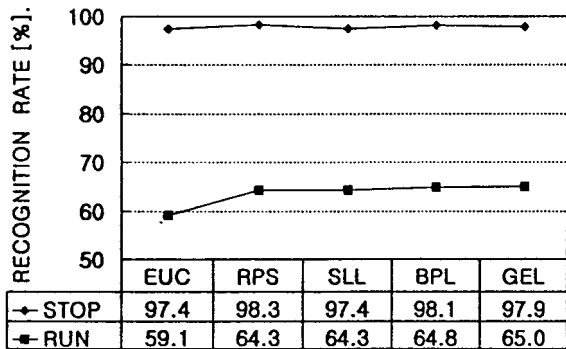


그림 3. 거리측정 방법에 따른 인식률 [%]  
Fig. 3. The recognition rates according to distance measure method

5.3 소음 처리 2 : High Pass Filtering

자동차 주행 소음은 1 kHz 이하의 저주파 대역에 주로 분포되어 있다. 따라서, 고역 통과 필터(HPF)를 사용하여 소음을 제거할 수 있다. 그러나 1 kHz 이하에는 음성 신호의 중요한 정보도 포함되어 있으므로, 차단 주파

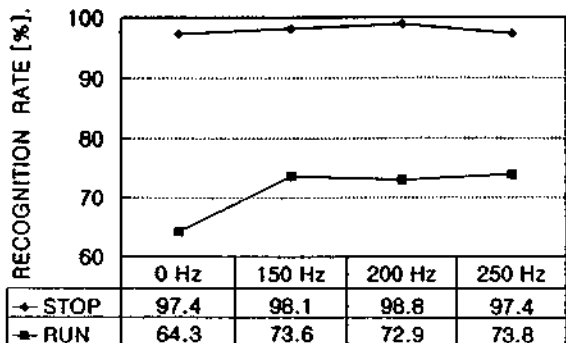


그림 4. HPF의 차단 주파수에 따른 인식률 [%] (SLL 사용 중)  
Fig. 4. The recognition rates according to the cut-off frequency of the HPF.

수(cutoff frequency)의 선택에 주의가 필요하다. HPF의 차단 주파수에 따른 인식 성능을 그림 4에 나타내었다.

그림 4의 인식률은 HPF와 함께 사용했을 경우 가장 좋은 성능을 보인 SLL 거리측정 방법을 사용한 인식률이다. HPF의 차단 주파수가 250 Hz 이상이면 정지 중의 인식률이 저하되기 시작했다. 이것은 음성 신호에 손상이 가해지는 것을 의미하므로, 이보다 낮은 200 Hz를 적절한 차단 주파수로 선정하였다.

5.4 소음 처리 3 : Smoothing

학습 과정에서 플로어 스무딩이 사용된 인식 시스템에 제안된 관찰 확률 스무딩을 추가하여 인식 성능을 측정하였다. 제안된 관찰 확률 스무딩 기법에는 스무딩할 관찰 확률의 수를 결정하는  $K$ , 스무딩을 위한 삼각형의 꼭지점 높이를 결정하는  $\alpha$ , 스무딩 삼각형의 기울기를 결정하는  $\beta$ 의 세 가지 상수값을 주어야 한다. 이들 세 가지 상수의 최적의 값을 구하기 위해 하나씩 변화시키면서 인식 성능을 측정하였다.

먼저  $K$ 의 최적값을 구하기 위한 인식 실험을 수행하였으며, 그 결과를 그림 5에 나타내었다.

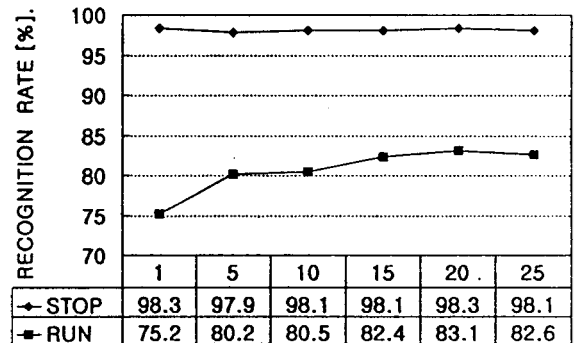


그림 5. 스무딩할 관찰 확률 수  $K$ 의 변화에 따른 인식률 [%] ( $\alpha=0.3, \beta=0.01$ )  
Fig. 5. The recognition rates according to  $K$ , the number of the observation probability to smooth.

$K$ 의 값을 변화시켜도 정지 중 인식률의 저하는 거의 없었으며, 주행 중 인식률은  $K$ 의 값이 20일 경우 가장 높은 결과를 보였다. 따라서  $K$ 의 최적값은 20이라고 판단하였다.

그림 6은  $\alpha$ 의 값을 변화시키면서 얻은 인식률이다.

$K$ 값 변화 실험의 결과와 유사하게 정지 중 인식률의 저하는 거의 없었으며, 주행 중 인식률은  $\alpha$ 의 값이 0.3인 경우 가장 높았다.

마지막으로  $\beta$ 의 값을 변화시키면서 측정한 인식률을 그림 7에 나타내었다.

$\beta$ 의 값은 0.01인 경우 가장 좋은 인식 성능을 보였다.

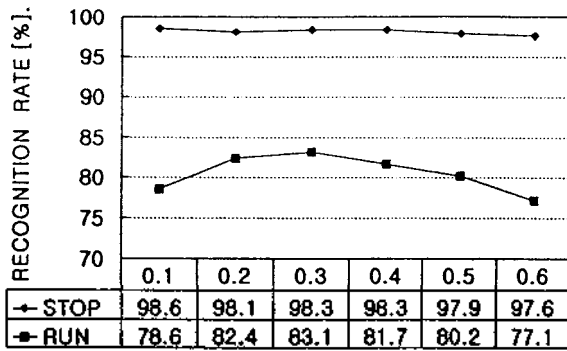


그림 6. 스무딩 삼각형의 꼭지점 높이를 결정하는  $\alpha$ 의 변화에 따른 인식률 [%] ( $K = 20, \beta = 0.01$ )

Fig. 6. The recognition rates according to  $\alpha$ , the height of the smoothing triangle.

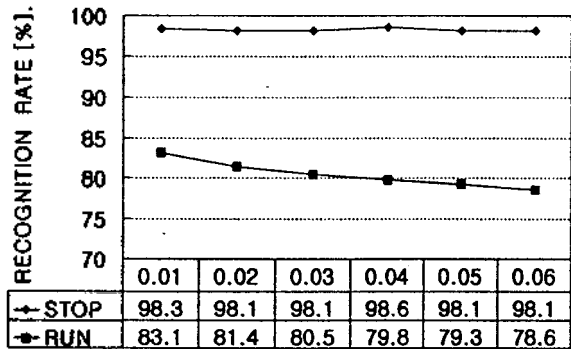


그림 7. 스무딩 삼각형의 기울기를 결정하는  $\beta$ 의 변화에 따른 인식률 [%] ( $K = 20, \alpha = 0.3$ )

Fig. 7. The recognition rates according to  $\beta$ , the slope of the smoothing triangle.

특징벡터의 거리 측정에 SLL을 사용하고, 200 Hz의 차단 주파수를 갖는 HPF를 적용시킨 인식 시스템의 인식률은 정지 중 98.8%, 주행 중 72.9%인데 비해, 제안된 관찰 확률 스무딩 기법을 추가한 시스템의 인식률은 정지 중 98.3%, 주행 중 83.1%로 정지 중의 인식률은 0.5% 낮아졌으나, 주행 중 인식률은 10.2%의 성능 향상을 보였다.

### 5.5 소음 처리 4 : Spectral Subtraction

최종적으로 스펙트럴 차감법[11]을 인식 시스템에 추가하였다. 기본 인식 시스템에서부터 소음 처리 기법을 하나씩 추가하여 마지막 스펙트럴 차감법까지 모든 소음 처리 기법이 결합된 인식 시스템의 인식 성능을 그림 8에 나타내었다. 스펙트럴 차감법은 일반적으로 소음 제거에 좋은 성능을 보이지만 음성 신호에도 왜곡을 준다는 단점을 갖고 있다. 그림 8에서도 이와 같은 결과를 보이고 있으며, 주행 중의 인식률은 향상되었으나 음성 신호 왜곡에 의해 정지 중의 인식률이 상당히 저하되었다.

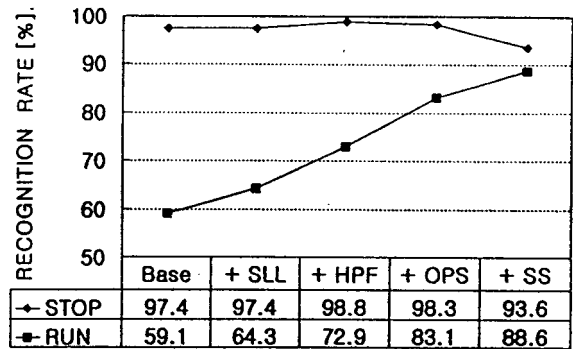


그림 8. 소음 처리 기법 추가에 따른 인식률 향상 [%]

Fig. 8. Enhancement of the recognition rates by the noise processing procedures.

### 5.6 최종 인식 시스템

스펙트럴 차감법이 정지 중의 인식률을 크게 저하시키는 문제를 해결하기 위해 비선형 스펙트럴 차감법[12]을 응용한 방법을 인식 시스템에 적용하였다. 적용된 방법은 입력되는 신호의 처음 10 프레임울 음성이 없는 묵음 구간이라 가정하고 평균 에너지를 구한다. 이 때 구해진 평균 에너지가 정해진 임계값(threshold)보다 크면 소음이 많은 상황이라 판단하여 스펙트럴 차감법을 수행하고, 임계값보다 작으면 소음이 적은 상황이라 판단하여 스펙트럴 차감법을 수행하지 않는 방법이다.

자동차에서의 단독 숫자음 및 명령어 인식을 위한 인식 시스템은 최종적으로 다음 그림 9와 같이 구성되었으며, 그 성능은 표 3과 같다.

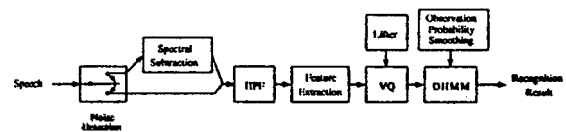


그림 9. 자동차에서의 단독 숫자음 및 명령어 인식을 위한 인식 시스템 구성

Fig. 9. The block diagram of the recognition system for isolated digit and command in car environment.

표 3. 최종 인식 시스템의 인식률 [%]

Table 3. The recognition rates of final recognition system.

	STOP	RUN
RECOGNITION RATE [%]	98.3	88.6

스펙트럴 차감법을 선택적으로 적용한 결과, 주행 중 인식에서의 인식률 향상은 물론 정지 중 인식에서도 음성 신호 왜곡에 의한 인식률 저하를 막을 수 있었다. 선택적인 스펙트럴 차감법의 적용을 위해서는 에너지 임계값이

필요하며, 이는 실험을 통해 구해야 한다. 본 논문의 인식 실험에 사용된 데이터는 자동차 주행 중과 정지 중이라는 차이가 매우 큰 상황에서 수집된 데이터이므로 정확한 임계값을 비교적 쉽게 찾을 수 있었다. 그러나, 여러 가지 다양한 상황에서 사용될 인식 시스템을 위한 임계값의 경우에는 매우 세심한 주의를 기울여 결정해야 하며, 스펙트럴 차감법의 선택적 적용보다는 일반적인 비선형 스펙트럴 차감법의 성능이 더 우수할 것으로 예상된다.

## VI. 결론

본 논문에서는 자동차 환경에서 DHMM을 이용한 단독 숫자음 및 명령어 인식에 대해 기술하였다. 자동차 소음에 대처하기 위한 방안으로 관찰 확률 스무딩 방법을 제안하였으며, 기존의 리프터, 고역 통과 필터, 스펙트럴 차감법 등을 적용하여, 자동차 주행 소음에 대한 효과적인 대처 기법을 제시하였다.

제안된 관찰 확률 스무딩 방법은 소음으로 인해 입력 음성의 특징벡터가 알맞은 코드워드로 정확히 양자화되지 못하는 경우에 대비해 근접한 코드워드들의 관찰 확률값을 높여주는 기법으로, 자동차 주행 중과 정지 중의 두 가지 상황에 대한 한국어 단독 숫자음 및 명령어 인식 실험 결과, 소음이 없는 경우에는 98.8%에서 98.3%로 0.5%의 인식을 저하가 있었으나, 소음이 존재하는 경우 72.9%에서 83.1%로 10.2%의 높은 인식 성능 향상을 얻을 수 있었다.

이와 함께 본 논문에서 제시한 쉐프스트럴 영역에서의 가중함수인 SLL의 사용, 200 Hz의 고역 통과 필터링, 스펙트럴 차감법의 적용 등은 자동차 주행 소음에 대처하기 위한 효과적인 기법으로 판단된다. 이 중 고역 통과 필터링은 음성 신호에 손상을 주지 않으므로 정지 중 인식 성능이 저하되지 않았지만, 자동차 주행 소음의 제거 정도는 스펙트럴 차감법에 비해 다소 떨어졌다. 반면, 스펙트럴 차감법은 소음 제거 정도는 크지만 음성 신호에 손상을 주어 정지 중 인식 성능이 크게 저하되었다. 따라서, 본 논문에서는 이 두 가지 방법을 동시에 사용하면, 소음의 존재 여부에 따라 스펙트럴 차감법을 선택적으로 적용하는 방법을 채택하였다. 그 결과 정지 중의 인식 실험에서 스펙트럴 차감법의 음성 신호 왜곡에 의한 인식 성능 저하를 방지하는 동시에 주행 중 인식 성능을 향상시킬 수 있었다. 최종적으로 얻은 인식률은, 정지 중의 경우 97.4%의 기본 인식률로부터 소음 처리 과정들을 추가한 후 98.3%를 얻었으며, 주행 중의 경우 59.1%에서 88.6%로 29.5%의 인식을 향상을 얻었다.

## 참고 문헌

1. S. E. Levinson, L. R. Rabiner, and M. M. Sondhi, "An Introduction to the Application of the Theory of Probabilistic Functions on a Markov Process to Automatic Speech Recognition," *The Bell System Technical Journal*, vol. 62, no. 4, pp. 1035-1074, Apr. 1983.
2. M. Cravero, L. Fissore, R. Pieraccini, and C. Scagliola, "Syntax Driven Recognition of Connected Words by Markov Models," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, vol. 3, pp. 35.5.1-35.5.4, San Diego, California, USA, Apr. 1984.
3. R. M. Schwartz, J. Klovstad, J. Makhoul, and J. Sorensen, "Improved Hidden Markov Modeling of Phonemes for Continuous Speech Recognition," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, vol. 3, pp. 35.6.1-35.6.4, San Diego, California, USA, Apr. 1984.
4. K. Sugawara, M. Nishimura, K. Toshioka, M. Okochi, and T. Kaneko, "Isolated Word Recognition Using Hidden Markov Models," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, pp. 1-4, Tampa, Florida, USA, Apr. 1985.
5. K. F. Lee, and H. W. Hon, "Large-Vocabulary Speaker-Independent Continuous Speech Recognition," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, pp. 123-126, New York, USA, Apr. 1988.
6. N. Nocerino, F. K. Soong, L. R. Rabiner, and D. H. Klatt, "Comparative Study of Several Distance Measure for Speech Recognition," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, pp. 25-28, Tampa, Florida, USA, Apr. 1985.
7. B. A. Hanson and H. Wakita, "Spectral Slope Distance Measure with Linear Prediction Analysis for Word Recognition in Noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 7, pp. 968-973, July 1987.
8. F. Itakura and T. Umezaki, "Distance Measure for Speech Recognition Based on the Smoothed Group Delay Spectrum," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, pp. 25-28, Dallas, Texas, USA, Apr. 1987.
9. B. H. Juang, L. R. Rabiner, and J. G. Wilpon, "On the Use of Bandpass Liftering in Speech Recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 7, pp. 947-954, July 1987.
10. J. Junqua and H. Wakita, "A Comparative Study of Cepstral Lifters and Distance Measures for All Pole Models of Speech in Noise," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, pp. 476-479, Glasgow, Scotland, May 1989.
11. S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 27, no. 2, pp. 113-120, April 1979.
12. P. Lockwood and J. Boudy, "Experiments with a Nonlinear Spectral Subtractor (NSS), Hidden Markov Models and Projection, for Robust Speech Recognition in Cars," *Speech Communication*, vol. 11, no. 2-3, pp. 215-228, Dec. 1992.

### ▲양 태 영(Tae Young Yang)

한국음향학회지 1996년 15권 2호 참조

현재 : 연세대학교 대학원 전자공학과 박사과정

1. S. E. Levinson, L. R. Rabiner, and M. M. Sondhi, "An Introduction to the Application of the Theory of Probabilistic Functions on a Markov Process to Automatic Speech Recognition," *The Bell System Technical Journal*, vol. 62,

▲신 원 호(Won Ho Shin)

한국음향학회지 1996년 15권 2호 참조

현재: 연세대학교 대학원 전자공학과 박사과정

▲김 지 성(Ji Sung Kim)

한국음향학회지 1998년 17권 6호 참조

1996년 9월~1998년 8월: 연세대학교 대학원 전자공학과  
(공학석사)

현재: (주) 로커스(Locus)

▲안 등 순(Dong Soon Ahn)

한국음향학회지 1995년 14권 4호 참조

현재: 목포대학교 컴퓨터공학과 교수

▲이 충 응(Chung Yong Lee)

한국음향학회지 1998년 17권 6호 참조

현재: 연세대학교 기계전자공학부 조교수

▲윤 대 희(Dae Hee Youn)

한국음향학회지 1994년 13권 1호 참조

현재: 연세대학교 기계전자공학부 교수

▲차 일 환(Il Whan Cha)

한국음향학회지 1995년 14권 4호 참조

현재: 연세대학교 기계전자공학부 교수