

스펙트럼사상을 이용한 잡음환경음성인식에 관한 연구

이 기 영

관동대학교 전자통신공학과

A Study on Speech Recognition in Noise Environment Using Spectral Mapping

Ki Young Lee

Dept. Electronic Comm. Eng., Kwan Dong Univ.

요 약

정적지도 화자적용기법에서 스펙트럼 거리에 의존하는 비선형적인 스펙트럼사상법을 이용하여 잡음환경에서의 음성인식방법에 관하여 연구한 결과, Top2에서 인식율의 향상을 얻어 그 유효성을 확인하였다. 본연구에서는 스펙트럼 거리에 의존하지 않는 선형 스펙트럼 사상법을 제시하고 그에 의한 잡음환경의 음성인식결과를 비선형적인 스펙트럼 사상법에 의한 결과와 비교하였다. 그결과, 인식율이 개선되었을 뿐 만아니라, Top1에서도 인식율이 향상되어 선형 스펙트럼사상법이 잡음환경음성인식방법으로 효과적인 방법임을 확인하였다.

1. 서 론

최근의 실용화되고 있는 음성인식장치는 아직까지 특정화자용 인식장치를 중심으로 개발되고 있으며, 사용자의 응용범위는 이미 구성된 인식장치의 표준화자가 발생한 음성의 범위에서 크게 벗어나지 못하고 있다. 단일, 지정된 특정화자가 아닌 다른 사용자가 그인식장치를 사용할 경우 인식성능은 매우 저하되어 버리기 때문이다. 이를 해결하기 위한 방법으로 벡터양자화를 이용하여 정적지도 화자적용방법[1,2]을 제시한 바 있으며, Feature Map을 이용하여 동적특자 화자적용방법[3]을 제시하여 아들에 대한 연구가 진행되고있다.

또한, 음성인식장치의 활용범위가 넓어질에 따라 잡음환경에도 강한 음성인식장치의 필요성이 강조되고 있는 것은 외부 기계로부터 음성에 부가되는 잡음, 외부환경이나 녹음실의 잡음

및 마이크나 화자의 성도에 의한 스펙트럼 왜곡등이 음성인식장치의 인식성능을 저하시키고 있기 때문이다.

Acerof[4]등의 연구결과에 따르면 불특정화자용으로 설계된 음성인식장치의 마이크나 외부환경의 잡음을 달리 했을때 인식성능이 매우 저하된다는 사실을 확인하였다. Bol[5]은 DFT계수에 스펙트럼차감법을 도입하여 음성중의 잡음제거를 시도하였으며, Stockham[6]등은 잡음보상방법으로 스펙트럼 동기회로를 사용하였다. Van Comperolle[7]는 스펙트럼차감법과 동기회로를 모두 사용하여 비교적 우수한 결과를 얻었으나, 이들 방법은 잡음의 스펙트럼 추정치의 독립성을 가정하고 이루어졌으므로 잡음의 제한된 범위에서 적용가능하다.

본인[8]은 정적지도 화자적용기법에서 스펙트럼 거리에 의존하는 비선형적인 스펙트럼사상법을 이용하여 잡음환경에서의 음성인식방법에 관하여 연구한 결과, Top2에서 인식율의 향상을 얻어 그 유효성을 확인한 바 있다. 그러나, 스펙트럼 거리에 의존하는 비선형적 스펙트럼사상을 이용하면 잡음환경의 음성과 비잡음환경에서의 음성 중 자음부분의 스펙트럼이 유사하기 때문에 음성에 포함된 잡음을 배제하기 위한 스펙트럼사상이 효과적으로 이루어 지기 어렵다.

따라서, 본연구에서는 스펙트럼 거리에 의존하지 않는 선형 스펙트럼 사상법을 제시하고 그에 의한 잡음환경의 음성인식결과를 비선형적인 스펙트럼 사상법에 의한 결과와 비교하였다. 그결과, 인식율이 개선되었을 뿐 만아니라, Top1에서도 인식율이 향상되어 선형 스펙트럼사상법이 잡음환경음성인식방법으로 효과적인 방법임을 확인하였다.

2. 스펙트럼 사상법

스펙트럼의 사상은 잡음환경의 음성신호에 대한 스펙트럼의 영역을 비잡음환경의 스펙트럼 영역으로 사상하여 얻을 수 있다.

이러한 사상은 서로 대응하는 두 코드벡터 시계열을 근간으로 이루어 지므로 먼저 벡터 양자화된 동일단어로 부터 대응하는 코드벡터의 histogram을 얻은 후, 이 histogram을 가중치로 하여 비잡음환경의 B 코드벡터들을 선형결합하여 사상코드북을 얻는다. 다음은 선형 스펙트럼사상법으로 잡음환경의 A 코드북으로부터 비잡음환경의 B 코드북으로의 사상코드북을 얻는 과정이다.

- (i) 비잡음환경과 잡음환경의 학습단어에 대한 코드북을 생성한다.
- (ii) 잡음환경의 코드북과 비잡음환경의 코드북사이의 거리 행렬을 구한다.
- (iii) 선형적으로 대응하는 코드벡터들의 히스토그램을 구한다.
- (iv) 다음 식에 의해 사상코드북을 구한다.

$$V_i(A \rightarrow B) = \sum_{j=1}^L h_{ij} V_j(B) / \sum_{j=1}^L h_{ij} \quad (1)$$

여기서,

$V_i(A)$ 는 잡음환경의 A 코드북의 i번째 코드벡터.

$V_j(B)$ 는 무잡음환경의 B 코드북에서의 j번째 코드벡터.

$V_i(A \rightarrow B)$ 는 사상코드북의 i번째 코드벡터.

h_{ij} 는 $V_i(A)$ 와 $V_j(B)$ 의 histogram.

L 은 코드북의 크기

이다.

이와같이 스펙트럼사상을 위한 사상코드북을 작성하는 학습 과정의 구성도를 그림 1에 나타내었다. 또한 스펙트럼사상을 행한 후 잡음환경의 음성을 인식하기 위하여 먼저 사상코드북과 표준코드북과의 거리행렬을 구하고 벡터양자화와 DTW를 이용한다. 이러한 인식과정의 구성도는 그림 2에 나타내었다.

3. 실험 결과 및 고찰

3-1. 실험 데이터

실험에서 사용된 음성은, 방음처리가 되지 않은 실내에서 남성 2인의 화자에 의해 자연스럽게 5 회씩 발생한 한국어 숫자 음들이다. 잡음환경의 음성은 이상의 한국어 숫자음에 SNR이 20[db]이 되도록 백색잡음을 섞어 사용하였다.

이 중에서 표준화자와 잡음환경의 학습단어는 각 화자가 1 회 발생한 것을 사용하였으며, 학습과정에서 사용하는 DTW는 기울기가 1 이고 대칭 형태인 경로를 이용하였다. 1, 2, 3 회에 발생한 것으로는 멀티템플릿으로 이용하기 위해 세 개의 표준템플릿을 작성하였으며, 이러한 음성들을 차단주파수가 3.4 kHz인 저역통과 필터에 통과시킨 후, 샘플링 주파수가 10 kHz인 AD 컨버터 (12-bit resolution) 로 샘플링하였다.

음성신호의 프레임 구간은 20.0 ms (200 샘플)로 하였으며 이동구간은 10.0 ms (100 샘플)로 하여 50%가 중첩되게 하였다. 벡터 양자화를 위한 프레임별 특징벡터는 14차 선형예측 계수벡터와 자기상관계수벡터로 하였으며, 각 음성의 코드북은

LBG 알고리즘에 의하여 생성하였다.

3-2. 인식결과와 비교 및 고찰

음성인식에 사용한 거리측정방법은 잡음환경에서 알려진 log likelihood ratio 를 사용하였으며, 벡터양자화와 DTW를 이용하여 음성을 인식하였다. 다음 표 1 에는 4가지 상태의 음성에 대한 코드북, 즉, 무잡음환경의 코드북, 잡음환경의 코드북, 비선형 스펙트럼 사상을 통한 사상코드북, 및 선형 스펙트럼을 통한 사상코드북을 각각 이용하여 인식율을 비교하였다. 이로부터 비선형 스펙트럼 사상에 의해 얻은 Top 1 과 Top 2 에서의 평균 인식율이 77.0[%]에서 79.5[%]로 개선되었을 뿐만 아니라, Top 1 에서도 평균인식율을 향상시켰으며, 코드북의 크기가 16일 때 92[%]의 인식율을 얻으므로써 잡음환경음성인식방법으로서 효과적임을 확인하였다.

잡음환경의 음성에 대해 사상코드북을 이용하지 않을 경우에는 코드북들에 대한 평균 인식율 72[%] 이고, 사상코드북을 이용할 경우에는 모든 코드북에 대한 평균 인식율 86.3[%]을 얻으므로써 잡음환경의 음성에 대한 오인식율을 약 1/2로 감소됨을 확인할 수 있었다. 여기서, Top 2 에 해당하는 음성은 대부분이 유사단어이며, '구', '공', '오' 사이에서나 '일', '이', '칠' 및 '삼', '사' 등에서 Top 2 에 주로 발생하였음을 확인하였다. 이

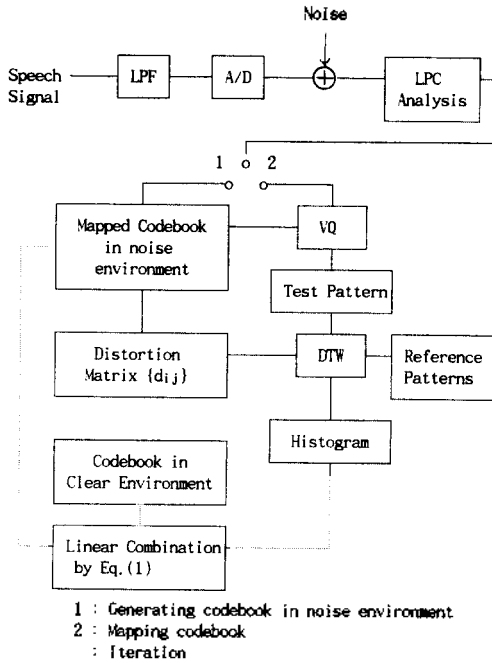


그림 1. 잡음환경에서 사상코드북을 작성하는 학습과정의 구성도
 Fig. 1. Block diagram of training for making a mapped codebook in noise environment

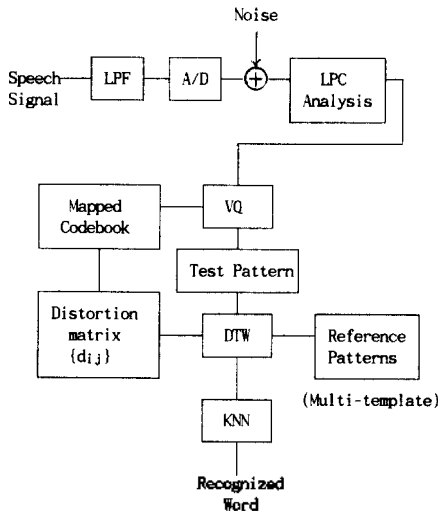


그림 2. 사상코드북에 의한 잡음환경음성인식의 구성도
 Fig. 2. Block diagram of word recognition in noise environment using a mapped codebook

단어들의 유사성은 주로 발성길이 가 긴 모음부분에 기인한 것으로 이를 개선하기 위해서는 자음 특히 파열음 'ㄱ'과 비음 'ㄴ' 등의 구별에 뛰어난 음성의 특징파라미터나 거리측정방법을 사용하면, Top 1 에서도 인식율이 더욱 개선될 것으로 사료된다.

표 1. 잡음환경의 인식율의 비교
 Table 1. Comparison of recognition rates in noise environment

CBK SIZE	8		16		32		Avg.	
	TOP 1	TOP 2	TOP 1	TOP 2	TOP 1	TOP 2	TOP 1	TOP 2
CLEAR	90	98	94	98	92	98	92	98
NOISE	54	56	68	68	74	74	72	73
N.MAPPED	48	76	86	95	70	88	68	86
L.MAPPED	53	76	92	95	76	85	74	85

N.MAPPED : Nonlinear spectral mapping

L.MAPPED : Linear spectral mapping

4. 결 론

본 연구에서는 사상코드북을 이용한 잡음환경의 음성인식방법에 관하여 연구하였으며 사상코드북의 작성방법으로 정칙지도 화자적용기법의 스펙트럼상상법을 이용하였고, 인식방법으로는 벡터양자화와 DTW를 이용하였다. 이를 평가하기 위하여 음성에 20[db]의 백색잡음을 섞어 잡음환경의 음성으로 음성인식실험을 수행한 결과, 평균인식율을 개선시킬 수 있었으며, TOP 2 에서는 오인식율을 1/2로 감소시키므로써 사상코드북에 의한 잡음환경음성인식의 유효성을 확인하였다.

<<참고 문헌>>

1. K. Shikano, K. F. Lee, R. Reddy, "Speaker Adaptation Through Vector Quantization," Proc. ICASSP 86, pp. 49. 5, 1986
2. 이기영, 최갑석, "사상맵퍼블입수에 의한 화자적용 단어인식," 한국음향학회지 제11권 제3호 pp. 40-52, 1992
3. R.M. Stern, "Dynamic Speaker Adaptation for Feature-Based Isolated Word Recognition," IEEE Trans. on ASSP, pp. 751-763, 1987
4. A. Acero, R.M. Stern, "Environmental Robustness in Automatic

- Speech Recognition," Proc. ICASSP 90, pp. 849-852, 1990
5. J.E. Porter, S.F. Boll, "Optimal Estimators for Spectra Restoration of Noisy Speech," Proc. ICASSP84, pp. 18A. 2.1-4. 1984
 6. T.G. Stockham, et al., "Blind Deconvolution Through Digital Signal Processing," Proc. IEEE, Vol. 63, pp. 678-692, 1975
 7. D. Van Compenolle, "Noise Adaptation in a HMM Speech Recognition System," Computer, Speech and Language, Vol. 3, pp. 151-167, 1989
 8. 김원구, 외3인, "잡음환경에서 음성인식을 위한 신호처리," 한국음향학회지 제11권 제2호 pp. 73-84, 1992
 9. 이기영, 외2인, "사상코드북을 이용한 잡음환경음성인식에 관한 연구," 1993년도 추계종합학술발표회, 한국통신학회, 1993