전화망을 통한 핵심어 검출 시스템에서의 채널왜곡 모상방법의 성능비교

이 교혁, 김 형순
부산대학교 전자공학과

Performance Comparision of Channel Distortion Compensation Techniques in Keyword Spotting System over the Telephone Network

Kyo Hyuk Lee, Hyung Soon Kim
Pusan National University, Dept. of Electronics Engineering

요 약

본 논문에서는 핵심어 검출(keyword spotting) 시스템에서의 채널왜곡에 대한 보상방법들의 성능을 비교하였다. 훈련용 음성과 인식실험용 음성은 서로 다른 환경에서 수집되었으며, 특별한 언급이 없을 음성으로는 전화망을 통한 음성 데이터를 이용하였다. 전화망을 통한 음성인식에서는 채널왜곡과 부가잡음에 의해서 음성신호에 왜곡이 생기므로 이들에 대한 적절한 보상이 필요하다. 본 논문에서는 채널왜곡 보상을 위한 처리 방법으로 널리 사용되고 있는 global cepstral mean subtraction(GCMS), local cepstral mean subtraction(LCMS) 그리고 RASTA processing을 적용하였다. 그리고 인식성능의 개선을 위해 이들 방법에 likelihood ratio scoring에 의한 후처리 과정을 적용하였다. 인식실험 결과 이들 방법 모두 채널왜곡 보상을 하지 않았을 경우와 비교하여 더 좋은 인식성능을 얻을 수 있었으며, 그 중 후처리를 적용한 LCMS 방법이 가장 우수한 성능을 나타내었다.

1. 서 론

전화망을 통한 음성인식은 음성인식의 매우 유망한 응용분야 중 하나이다. 그러나, 전화망을 통한 음성인식은 연구실 환경에서의 고음질 음성의 인식에서는 고려할 필요가 없었던 많은 문제점들을 지닌다. 그 대표적인 예로 채널대역폭 제한 및 handset 마이크로폰 특성에 의한 왜곡과 배경잡음의 증가 등을 들 수 있다. 따라서, 전화망을 통한 음성인식의 성공함상을 위해서는 채널왜곡과 배경잡음에 대한 효과적인 보상방법이 필수적으로 요구된다.

일반적으로 전화망에 의한 채널왜곡은 음성신호를 linear filtering 하는 것으로 모델링할 수 있다. 이 때 filter가 time-invariant 라고 가정한다면 채널왜곡은 음성의 cepstrum에 bias를 가해주는 것으로 해석될 수 있다. 따라서 이러한 bias를 제거해 줌으로써 인식성능의 향상을 기대할 수 있다. 이러한 채널왜곡 보상 방법으로는 global cepstral mean subtraction(GCMS), local cepstral mean subtraction(LCMS)[10], cepstral bias removal(CBR)[6], codeword-dependent cepstral normalization(CDCN)[8], RASTA(RelAtive SpecTrAl) 처리[4] 등이 있다. 이들 방법 중에서 GCMS, LCMS 및 RASTA 처리 방법이 계산량에 비해 우수한 성능을 나타내는 것으로 알려져 있다.

본 논문에서는 음소 HMM에 의한 핵심어 검출 시스템을 구성하고, 이를 전화망을 통한 음성인식에 적용하였다. 전화망을 통한 음성인식을 위해서는 훈련용 음성 데이터도 전화망을 통해 수집하는 것이 바람직하다. 그러나 본 논문에서는 이미 구성된 고음질의 음성 데이터 베이스를 전화망 음성인식을 위한 훈련용으로 사용할 수 있는가 여부를 확인하기 위해 훈련용 데이터는 clean speech를 그대로 사용하였으며, 인식 실험용 데이터은 Dialogic 사의 D/41D VMS board를 이용해서 전화망을 통해 수집된 데이터를 사용하였다.

채널왜곡 보상을 위한 처리방법으로는 GCMS, LCMS와 RASTA 처리방법을 사용하였으며, 이들 방법 중 LCMS 방법이 가장 우수한 성능을 나타내었다.

2. Baseline 핵심어 검출 시스템

본 논문에서 사용한 핵심어 검출 시스템의 전체 구성도를 그림 1에 나타내었다. 먼저 입력음성이 들어오면 음성특징 파라메터를 추출하여 훈련에 의해 만들어진 핵심어 모델, 비핵심어 모델 그리고 묵음 모델을 이용하여 Viterbi decoding 과정을 통해 핵심어를 검출한다. 본 논문에서는 음성특징 파라메터로 12차의 LPC 켑스트럼 및 12차 델타 켑스트럼을 사용하였다.

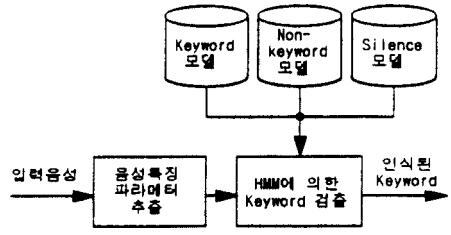

그림 1 핵심어 검출 시스템의 구성도

2.1 음소 HMM 모델

본 논문에서는 한국어에 대해 총 45개의 문맥 독립형(context-independent) 유사음소를 정의하고 이를 기준으로 triphone 모델을 구

성하였으며, 이 triphone 모델들을 발음사전에 따라 연결하여 핵심어 모델을 구성하였다. 본 논문에서 사용한 음소 HMM 모델의 topology는 그림 2와 같다. 음소 HMM 모델은 3개의 state와 6개의 transition을 가진며 transition에서 관찰벡터가 출력되는 구조를 가진다. 관찰확률분포는 그림에서 보는 바와 같이 B, M, E의 3가지 분포로 tying되어 있다.

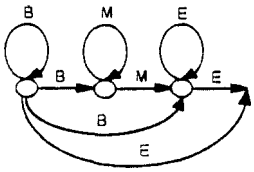

그림 2. 음소 HMM 구조

2.2 비핵심어 및 묵음 모델

핵심어 검출 시스템에서는 핵심어가 아닌 부분을 표현해 주기 위해 filler 모델, 즉 비핵심어 모델과 묵음 모델을 사용한다. 이러한 filler 모델을 어떻게 구성하는가에 따라 시스템의 인식성능은 크게 좌우된다. Filler 모델로는 핵심어 부분을 감식하지 않으면서 핵심어가 아닌 부분을 어느정도 표현해 줄 수 있는 모델이 필요하다[1].

본 논문에서 비핵심어 모델을 선택하기 위한 방법으로 통계적 방법에 의한 monophone clustering 방법을 사용하였다[5]. 먼저 45개의 monophone을 구성한 뒤 이들을 통계적 방법에 의해 clustering하여 6개의 비핵심어 모델을 구성하였다. Monophone clustering을 위한 distance measure는 다음과 같다.

$$D(p_i, p_j) = \sum_{d=1}^{S} D_d(p_i, p_j) \qquad (1)$$

여기서 P_i, P_j는 각각 i와 j번째 음소를 나타내고, S는 음소 모델의 distribution 수를 나타내며, $D_d(p_i, p_j)$는 두 음소의 d번째 distribution간의 distance로서 다음 식과 같이 주어진다.

$$D_d(p_i, p_j) = \frac{1}{V}\sum_{k=1}^{V}\frac{(\mu_{idk} - \mu_{jdk})^2}{\sqrt{\sigma_{idk} - \sigma_{jdk}}} \qquad (2)$$

여기서 V는 관찰 벡터의 차수를 나타내고, μ_{idk}와 σ^2_{idk}는 각각 i번째 음소의 d번째 distribution의 평균과 분산을 의미한다.

본 논문에서의 묵음 모델로는 10개의 state를 가지는 단일 모델을 사용하였다.

2.3 핵심어 검출을 위한 전체 HMM network 구조

핵심어 검출을 위한 전체 network은 핵심어 모델, 비핵심어 모델, 묵음 모델들을 사용하는 연결단어인식 알고리즘을 기반으로 하고 있다.

일반적으로 한 문장 내에는 임의의 핵심어 갯수가 올 수 있으므로 null grammar 형태가 가능하지만, 본 논문에서는 입력음성에 1개의 핵심어가 존재한다는 가정하에 그림 3과 같은 문법 구조를 가지는 network을 구성하였다.

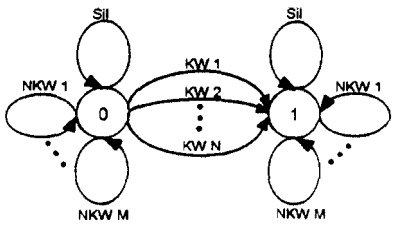

그림 3. 핵심어 검출을 위한 전체 HMM network

3. 채널왜곡 보상을 위한 방법

서론에서 언급하였듯이 일단 연결된 전화회선의 채널왜곡은 linear time invariant filter로 모델링할 수 있다. 따라서 이러한 filter의 영향은 cepstral domain에서 살펴볼 때 모음의 bias를 가해주는 것으로 해석된다.

$$\vec{z} = \vec{x} + \vec{h} \qquad (3)$$

여기서 $\vec{z}$는 관찰벡터, $\vec{h}$는 channel filter의 cepstral 벡터, 그리고 $\vec{x}$는 입력음성의 cepstral 벡터이다.

3.1 Global Cepstral Mean Subtraction(GCMS)

Global cepstral mean subtraction의 목적은 식(3)에서 bias 항목으로 나타나는 $\vec{h}$를 관찰벡터로부터 제거하는 것이다. 관찰벡터 $\vec{z}_{(i)}^{T}$에서 평균벡터는 다음과 같다.

$$\vec{m} = \frac{1}{T}\sum_{(i)}^{T}\vec{z}_i \qquad (4)$$

그리고 채널왜곡이 보상된 벡터는 다음과 같이 표현된다.

$$\vec{z}_i^{comp} = \vec{z}_i - \vec{m} \qquad (5)$$

이와 같이 채널왜곡이 보상된 벡터는 channel에 의한 bias에 영향을 받지 않는다. 그러나, CMS에 의한 보상방법은 channel bias 뿐만 아니라 입력음성 특징벡터의 평균까지도 함께 제거한다. CMS 방법은 입력음성 특징벡터의 평균제거에 의한 문제점보다는 채널왜곡에 의한 bias 제거의 긍정적 효과가 인식률에 미치는 영향이 더 크다는 것을 가정하고 있다. CMS 방법은 입력음성 전체에 대한 평균벡터를 구해야 하기 때문에 실시간 처리가 곤란하다는 단점을 지닌다.

3.2 Local Cepstral Mean Subtraction (LCMS)

Local cepstral mean subtraction 역시 global cepstral mean subtraction과 마찬가지로 식(3)에서 bias 항목으로 나타나는 $\vec{h}$를 관찰벡터로부터 제거하는 것을 목적으로 한다. GCMS에서는 bias를 관찰벡터 전체에 대한 평균을 bias로 정의하였으나 LCMS 방법에서

는 moving cepstral average를 벡터로 취한다. 관찰벡터 $\vec{x}_{t-1}^{T}$ 에서 moving cepstral average는 다음과 같다

$$\vec{m}_t = \frac{1}{T_L}\sum_{\tau=0}^{T_L-1}\vec{x}_{t-\tau}, \quad T_L < T \qquad (6)$$

그리고 채널왜곡이 보상된 벡터는 다음과 같이 표현된다

$$\vec{x}_t^{comp} = \vec{x}_t - \vec{m}_t \qquad (7)$$

LCMS 방법은 moving average를 취하기 때문에, channel distortion이 고정되어 있지 않고 변하더라도 그 영향을 제거해 줄 수 있다. LCMS 방법 역시 channel bias와 함께 입력음성 특징벡터의 평균을 제거한다. 그러나 Local CMS는 CMS와는 달리 실시간 처리가 가능하다.

3.3. RASTA (RelAtive SpecTrAl) processing

전화망에 의한 채널왜곡 특성은 대부분 스펙트럼 특성이 고정되어 있거나 변한다고 하더라도 음성신호의 변화 특성에 비해서 천천히 변화한다. 그리고 전화망에는 드라마또는 배경잡음도 음성신호의 스펙트럼 특성에 비해 천천히 변화하거나 급격히 변화한다고 가정할 수 있다. RASTA처리는 이와 같이 음성신호의 특성변화에 비해 매우 천천히, 또는 매우 빠르게 변화하는 채널왜곡 및 부가잡음의 영향을 제거하는 목적으로 입력특징 벡터들을 그림 4와 같은 주파수 특성을 가지는 bandpass filter로 filtering하는 방법이다[4].

RASTA filter의 transfer function은 다음과 같다

$$H(z) = 0.1z^{4}\frac{2 + z^{-1} - z^{-3} - 2z^{-4}}{1 - 0.98z^{-1}} \qquad (8)$$

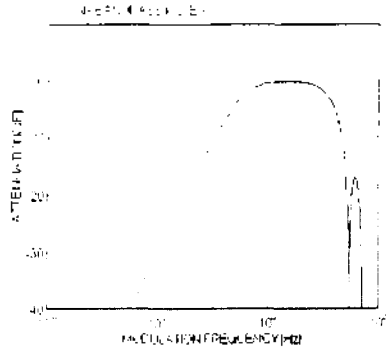

그림 4 RASTA filter의 주파수 특성

RASTA처리는 CMS 방법에 비해 실시간 처리에 유리하다.

4. Likelihood Ratio Scoring에 의한 후처리

후처리 과정은 핵심어검출 시스템에서의 오류를 감소시켜 그 성능을 보다 향상시키기 위한 것으로서, 이미 구해진 핵심어 후보들의 신뢰도를 판단하여 잘못 검출된 후보(false alarm)들을 효율적으로 제거하는 데에 주안점을 두고 있다.

이러한 후처리 방법들로는 신경회로망을 이용하는 방법, 음성 segment 모델을 이용하는 방법, 변별적 훈련과정을 사용하는 방법 그리고 likelihood ratio에 의한 방법 등이 있다. 후처리를 위한 여러 방법들 중 본 논문에서는 likelihood ratio에 의한 후처리를 수행하였다[9]. Likelihood ratio test는 핵심어 및 filler 모델의 likelihood 비를 이용하는 방법으로서, 그림 5에 도시된 바와 같이 두 가지의 HMM network를 병렬로 사용한다. 그 중 첫번째는 2장에서 설명한 핵심어 및 filler network으로서 그림 3에서 표현된 것과 같으며, 두번째는 핵심어 모델없이 filler 모델로만 구성된 network이다. 핵심어 및 filler network에 입력 문장으로부터 핵심어 후보를 검출해 내고 동시에 이 핵심어에 해당하는 음성구간에 대한 정보를 filler 모델만으로 구성된 network에 넘겨주면, 이 network은 해당 구간에서의 filler 모델의 likelihood 값을 계산한다. 그 다음 앞서 핵심어 및 filler network에서 핵심어 검출시에 계산된 likelihood 값과 filler network에서의 filler likelihood 값을 비교하여 그 비가 특정 임계치를 넘지 못하면 그 음성구간에서의 핵심어 검출의 신뢰도가 높지 못한 것으로 간주하여 검출된 핵심어를 기각시킨다. 그림 5에 likelihood ratio scoring에 의한 후처리 과정을 나타내었다.

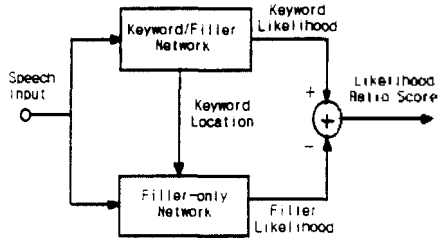

그림 5. Likelihood Ratio Scoring에 의한 후처리 과정

5. 데이터 베이스

전화망을 통한 핵심어 검출 시스템의 성능 평가를 위한 task domain으로는 자동전화교환 서비스를 선정하였으며 총무과, 자산관리과, 회계과, 내자과, 설비과 그리고 서울사무소의 6개 부서명을 인식대상 핵심어로 정하였다.

5.1 훈련용 데이터 베이스[8]

핵심어 모델을 위한 훈련용 데이터 베이스는 표준어를 구사하는 남성 35명이 6개의 부서명을 고립단어 형태로 각 1번씩 발음한 clean speech를 사용하였다. 이 데이터 베이스는 16KHz sampling, 16bit linear PCM으로 coding되어 있으며, 이것을 telephone network의 sampling rate에 맞추기 위해 8KHz로 downsampling하여 훈련에 사용하였다.

비핵심어 모델을 위한 데이터 베이스는 표준어를 쓰는 남성 22명이 음성학적으로 균형이 잡힌(phonetically balanced) 445단어를 각 1회씩 발음한 clean speech를 사용하였다. 이 데이터 역시 16KHz

sampling, 16bit linear PCM으로 coding되었던 것을 8KHz로 downsampling 하여 사용하였다.

5.2 인식용 데이터 베이스

인식용 데이터베이스는 부산 지방에 거주하는 남성 대학원생 15명이 6개 부서명에 대해 고립단어 및 문장형태로 각 1번씩 발음한 것을 사용하였다. 이 데이터 베이스는 실험실 환경에서 전화기를 사용하여 국선 전화망(PSTN) 및 구내교환기를 거쳐 8KHz sampling, 8bit μ-law PCM으로 녹음되었다. 이것을 다시 훈련용 데이터 베이스의 coding 방식에 맞추기 위해 16bit linear PCM으로 변환하였다. 음성 데이터 수집에는 Dialogic 사의 D/41D VMS board를 이용하였으며, 전화망을 통한 음성 데이터 수집 경로를 그림 6에 나타내었다.

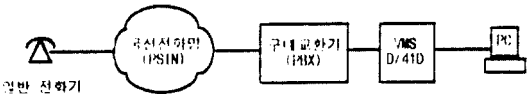

그림 6. 인식용 데이터 베이스의 수집 경로

6. 실험 결과 및 검토

훈련용 데이터의 녹음 상황과 같은 상황에서 녹음한 데이터에 대한 인식실험과 5.2절에서 설명한 전화망을 통한 데이터를 이용하여 global CMS(GCMS), local CMS(LCMS) 및 RASTA 방법을 적용한 인식실험을 하였다. 훈련용 데이터의 녹음 상황과 같은 상황에서 녹음한 데이터는 훈련용 데이터 수집에 참가하지 않은 표준어를 구사하는 남성 15명이 6개의 부서명을 고립단어 및 문장형태로 각 1번씩 발음한 clean speech로 구성되어 있다. 실험에 대한 결과를 표 1에 나타내었다.

표 1. Clean speech 및 전화망을 통한 음성을 이용한 인식결과

	모상 방법	고립단어	문장
Matched Condition	Baseline System	100%	92.00%
Mismatched Condition	No Processing	80.00%	54.44%
	RASTA	93.33%	66.67%
	GCMS	94.44%	76.67%
	LCMS(T_L=150msec)	94.44%	86.67%
	LCMS(T_L=300msec)	94.44%	83.33%
	LCMS(T_L=500msec)	91.11%	71.11%

표 1에서 RASTA 처리, GCMS 방법 및 LCMS 방법을 적용함으로써 채널왜곡에 대한 보상처리를 하지 않을 경우에 비해 성능이 크게 향상됨을 확인할 수 있다. 채널왜곡 보상을 하지 않은 경우와 비교해 볼 때, 문장형태의 경우 RASTA 처리방법, GCMS 방법에서의 오인식률이 각각 26.8% 및 48.9%만큼 감소되었으며, LCMS 방법에서 local length(T_L)가 150msec, 300msec, 500msec의 경우 각각 70.7%, 63.4% 및 38.6%만큼의 오인식률 감소를 얻었다. 고립단어 및 문장형태 모두에 대해 전반적으로 RASTA 처리가 가장 뛰어난 성능을 나타내었다. 고립단어에서는 GCMS 방법과 LCMS 방법이 성능 차이를 나타내지 않으나, 문장형태의 경우 LCMS 방법이 GCMS 방법에 비해 전반적으로 우수한 성능을 나타내고 있다. 이러한 결과는 문장형태의 음성을 GCMS 방법으로 인식할 경우 훈련시 추정한 평균과 인식시 추정한 평균이 일치하지 않기 때문에 GCMS 방법의 혼심어 모델이 문장에 포함된 혼심어를 표현하기에는 부적절하기 때문이라고 판단된다. LCMS 방법의 경우 local length가 길어질 수록 인식률이 감소함을 볼 수 있는데, 이는 참고문헌 [10]의 결과와 일치한다. 표 1에서 채널왜곡에 대한 보상 처리를 하지 않은 경우뿐만 아니라 GCMS, LCMS 방법 및 RASTA 처리를 하였을 경우에도 환경 차이(clean speech와 전화망을 통한 음성)에 의한 인식성능의 저하가 나타남을 알 수 있다. 이 결과는 GCMS, LCMS 방법 및 RASTA 처리 만으로는 훈련용 데이터와 인식용 데이터의 환경차이를 충분히 보상할 수 없음을 보여주는 것으로서, 여기에는 인식용 데이터에 상당한 크기의 배경잡음이 부가된 데에도 기인하는 바가 크다고 판단된다. 실제로 GCMS, LCMS 방법 및 RASTA처리는 채널왜곡의 보상에는 효과적이나 부가잡음의 제거에는 효과적이지 못하다.

전화망을 통한 인식용 데이터를 이용하여 GCMS, LCMS 및 RASTA 처리에 4절에서 설명한 후처리 과정을 적용시킨 결과를 표 2에 나타내었다. 후처리 과정에서는 전체 데이터 중 likelihood ratio가 낮은 것으로 약 8%를 기각하였다.

표 2. 전화망을 통한 음성을 이용한 후처리 결과
(후처리 과정을 통해 약 8%를 기각시킨 경우)

	고립단어	문장
No Processing	80.00%	50.00%
RASTA	96.51%	72.15%
GCMS	94.38%	78.95%
LCMS(T_L=150msec)	96.51%	86.08%
LCMS(T_L=300msec)	98.61%	89.47%
LCMS(T_L=500msec)	93.10%	75.64%

후처리를 하지 않았을 때와 비교하여 GCMS 방법, LCMS 방법 및 RASTA 처리의 경우 모두 추가적인 오인식률의 감소를 얻을 수 있었다. 특히 문장형태의 경우 RASTA처리방법, GCMS 방법에서 각각 12.18% 및 4.9% 그리고 LCMS 방법에서 local length(T_L)가 300msec, 500msec의 경우 각각 13.5% 및 9.7%만큼의 추가적인 오인식률의 감소가 이루어졌다.

표 1, 2에서 가장 좋은 인식률을 보인 후처리를 적용한 LCMS 방법(T_L=300msec)과 baseline system과의 인식률 차이는 고립단어와 문장형태에서 각각 4.49% 및 2.53%이다. 여기에는 인식용 데이터에 추가된 배경잡음에 의한 영향이 크게 작용한 것으로 판단된다. 이 결과는 궁극적으로는 훈련용 데이터 베이스도 전화망을 통해 수집해야하는 필요성을 보여준다.

훈련용 데이터는 표준어를 구사하는 화자에 의한 상옥체 음성으로서 발음이 명료한데 비해, 인식용 데이터는 부산지방의 방언을 쓰는 화자에 의해 대화체에 가깝게 자연스럽게 발음된 점도 인식 성능 저하의 한 요인으로 판단된다.

7. 결 론

본 논문에서는 핵심어 검출(keyword spotting) 시스템에서의 채널 왜곡에 대한 보상방법들의 성능을 비교하였다. 훈련용 음성과 인식용 음성의 환경차이를 기존의 채널왜곡 보상방법이 얼마만큼 해결할 수 있는가를 확인하기 위해 훈련용 음성으로는 clean speech를 사용하고 인식용 음성만 전화망을 통해 수집하였다. 채널왜곡의 보상방법으로는 RASTA처리, GCMS 방법 및 LCMS 방법을 적용하였다. 그리고 인식 성능의 개선을 위해 likelihood ratio scoring에 의한 후처리 과정을 적용하였다.

실험결과 문장형태의 경우 RASTA처리방법, GCMS방법에서의 오인식률이 각각 26.8% 및 48.9%만큼 감소되었으며, LCMS 방법에서 local length(T_L)가 15msec, 30msec, 50msec의 경우 각각 70.7%, 63.4% 및 36.6%만큼의 오인식률 감소를 얻었다. 후처리 과정을 통해 후처리를 하지 않았을 때와 비교하여 GCMS 방법, LCMS 방법 및 RASTA 처리의 경우 모두 추가적인 오인식률의 감소를 얻을 수 있었다. 특히 문장형태의 경우 RASTA처리방법, GCMS방법에서 각각 12.18% 및 4.9% 그리고 LCMS 방법에서 local length(T_L)가 300msec, 500msec의 경우 각각 13.5% 및 9.7%만큼의 추가적인 오인식률의 감소가 이루어졌다. 가장 좋은 인식률을 보인 방법은 후처리를 적용한 LCMS 방법(T_L=300msec)으로 그림단어와 문장형태에서 각각 95.51% 및 89.47%의 인식률을 나타내었다. 그러나 clean speech에 의한 인식결과와는 차이(그림단어와 문장형태에서 각각 4.49% 및 2.53%)를 보여 궁극적으로는 전화망을 통한 훈련용 데이터 수집의 필요성을 입증해 주었다.

참고문헌

[1] 김 광순, "Keyword Spotting 기술", 한국통신학회지, 제11권 9호, pp.57-65, 1994.

[2] C. H. Lee and L. R. Rabiner, "A Frame-Synchronous Network Search Algorithm for Connected Word Recognition", IEEE Trans. on ASSP, vol.37, no.11, pp.1649-1658, Nov. 1989.

[3] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", IEEE Trans. on ASSP, vol.37, no.6, pp795-804, June 1990.

[4] H. Hermansky, "RASTA Processing of Speech", IEEE Trans. on Speech and Audio Processing, vol.2, no.4, Oct. 1994.

[5] 이 황림, 김 형순 외, "음소 HMM을 이용한 Keyword Spotting 소시스템에서의 Non-Keyword 모델에 관한 연구", 제12회 음성통신 및 신호처리 워크샵, pp.83-87, 1995년 6월.

[6] M. G. Rahim and B. H. Juang, "Signal Bias Removal for Robust Telephone Based Speech Recognition in Adverse Environments," Proc. IEEE ICASSP, pp.445-448, 1994.

[7] A. Acero, *Acoustical Environmental Robustness in Automatic Speech Recognition*, Kluwer Academic Publishers, Boston, 1992.

[8] 이 영직, 류 준영, 김 상호, 황 규웅, "ETRI의 음성 데이터 베이스 구축 현황," 제 12회 음성통신 및 신호처리 워크샵 논문집, PP.265-267, 1995년 6월.

[9] R. C. Rose and D. B. Paul, "A hidden Markov model based keyword recognition system," in Proc. IEEE ICASSP, pp.129-132, 1990.

[10] A. E. Rosenberg, C. H. Lee and F. K. Soong, "Cepstral Channel Normalization Techniques for HMM-Based Speaker Verificatin," Proc. IEEE ICSLP, pp.1835-1838, 1994.