# 강인한 음성 인식을 위한 선형 로그 함수 기반의 MFCC 특징 표현 연구*

윤영선(한남대)

<차 례>



<Abstract>

## Representation of MFCC Feature Based on Linlog Function for Robust Speech Recognition

Young-Sun Yun

In previous study, the linlog(linear log) RASTA(J-RASTA) approach based on PLP was proposed to deal with both the channel effect and the additive noise. The extraction of PLP required generally more steps and computation than the extraction of widely used MFCC. Thus, in this paper, we apply the linlog function to the MFCC for investigating the possibility of simple compensation method that removes both distortion. With the experimental results, the proposed method shows the similar tendency to the linlog RASTA-PLP. When the $J$ value is set to 1e-6, the best ERR(Error Reduction Rate) of 33% is obtained. For applying the linlog function to the feature extraction process, the $J$ value plays a very important role in compensating the corruption. Thus, the study for the adaptive $J$ or noise dependent $J$ estimation is further required.

* Keywords: Robust speech recognition, Linear log, Linlog MFCC

* 이 논문은 2005년도 한남대학교 학술연구조성비 지원에 의하여 연구되었음

# 1. 서 론

음성인식이란 발성된 음성 파형을 텍스트로 변환하는 과정을 말한다. 사람의 말을 인식하여 텍스트로 변환하는 노력은 1930년대 중반이후로 오랫동안 진행되어오고 있으며, 그 결과 음성 인식 시스템의 성능은 일부분 사람의 능력에 근접하고 있다는 평가를 받는다. 그러나 널리 사용되는 음성 인식 접근 방법인 통계적 패턴 방법은 학습 환경과 평가 환경이 다를 경우 그 성능이 저하된다는 약점을 가지고 있다. 음성 인식 성능을 결정하는 환경적인 요소는 발성 화자, 발성 위치, 입력 장치, 발성 환경 등 여러 가지가 있다. 이러한 환경적 불일치를 제거하기 위한 연구는 화자적응, 음성 위치, 원거리 음성인식, 채널왜곡보정, 가산잡음제거 등 여러 분야에서 진행되고 있다. 이 중에서 본 논문은 채널 왜곡 보정과 가산 잡음 제거에 적용 가능한 방법을 제안하고 그 가능성을 살피고자 한다.

일반적으로 잡음은 전송되는 매체의 특성에 따라 음성 신호가 왜곡되는 채널 잡음과 음성이 입력되는 중간에 주위에서 음성 이외의 신호가 추가되는 가산 잡음(환경 잡음)으로 분류할 수 있다. 채널 잡음은 CMS(Cepstral Mean Subtraction 또는 Cepstral Mean Normalization)나 RASTA(RelAtive SpecTrAl)[2] 등과 같이 평균 값을 0으로하는 고대역 필터(high pass filter)를 통하여 제거할 수 있고, 가산 잡음은 잡음의 종류에 따라 모델 또는 자료기반 방식의 보상 방법 등을 사용한다. 또한 가산 잡음이 포함된 채널 왜곡을 보상하기 위하여 linlog(linear log) RASTA(또는 J-RASTA) 등의 방법이 제안되어 사용되고 있다[5].

RASTA 또는 linlog RASTA는 PLP(Perceptual Linear Predictive)[1] 특징에 적용되어 채널 잡음 또는 채널 잡음과 가산 잡음을 동시에 보상하기 위하여 제안된 기법이다. 본 연구에서는 PLP 특징 대신 널리 사용되는 MFCC(Mel-Frequency Cepstral Coefficients)에 linlog 함수를 적용하여 가산 잡음이 포함된 채널 잡음의 보상에 적용가능한지를 살펴보았다.

본 논문의 구성은 다음과 같다. 2장에서는 채널 보상 기법으로 널리 사용되는 CMS와 RASTA, 그리고 linlog RASTA에 대해 소개한다. 3장에서는 제안하는 linlog MFCC에 대해 설명한 후, 다음으로 linlog MFCC를 이용한 실험 및 그 결과를 정리한다. 마지막으로 5장에서는 요약 및 향후 연구에 대해 기술한다.

# 2. 채널 보상 기법

## 2.1. CMS

켑스트럼 고대역 필터링은 거의 계산량의 증가 없이 음성 인식의 강인성을 높

이는 방법으로 널리 사용되고 있으며, 채널 효과를 보이는 선형 필터와 가산 잡음을 동시에 보상하는 효과가 있다. 대표적인 방법으로는 CMS가 있으며, 이는 전체 발화 음성에서 전체 음성의 평균을 차감한다. 이 방법은 선형 필터를 제거하는 효과가 있으나 음성 평균을 마찬가지로 차감하기 때문에 음성 신호의 왜곡을 가져오게 된다.

$$\begin{aligned} \bar{c} &= \frac{1}{T}\sum_{t=1}^{T} c_t \\ \hat{c_t} &= c_t - \bar{c} \end{aligned} \tag{1}$$

즉, 음성 신호 $s_t$가 채널 필터 $f$의 영향을 받는다면 채널을 통과한 $x_t = f^* s_t$ 는 켑스트럼 영역에서 $c_t = \hat{f_t} + \hat{s_t}$의 형태로 모델링 된다. 여기에서 채널 필터 $\hat{f_t}$는 선형필터로서 시간의 영향을 받지 않는다고 하면, $\hat{f_t} \approx \hat{f}$로 근사될 수 있다. 결국 $\bar{c_t}$는 다음과 같은 특성을 갖는다.

$$\bar{c_t} = \frac{1}{T}\sum_{t=1}^{T} c_t = \hat{f} + \frac{1}{T}\sum_{t=1}^{T} \hat{s_t} \tag{2}$$

따라서 최종적으로 얻어지는 켑스트럼 $\hat{c_t}$는 다음의 식과 같이 채널 특성이 제거됨과 동시에 음성 신호가 작아지게 된다.

$$\begin{aligned} \hat{c_t} &= c_t - \bar{c_t} \\ &= (\hat{f} + \hat{s_t}) - (\hat{f} + \frac{1}{T}\sum_{t=1}^{T} \hat{s_t}) \\ &= \hat{s_t} - \frac{1}{T}\sum_{t=1}^{T} \hat{s_t} \end{aligned} \tag{3}$$

위 식에서 보는 것처럼 켑스트럼 평균 차감법을 통하여 얻어지는 켑스트럼 계수에서 평균 음성 신호가 제거됨으로써 음성 전반에 포함된 가산 잡음의 영향이 줄어든다.

### 2.2. RASTA와 linlog-RASTA

CMS와 더불어 널리 사용되는 고대역 필터 방법은 RASTA(RelAtive SpecTrAl) 필터링 방법으로, 말초 청각 시스템(peripheral auditory system)과 같이 음성 신호의

천이 부분(transient component)을 강조하기 위하여 고안되었다. 원 RASTA 필터[2]의 전달 함수는 IIR(Infinite Impulse Response) 필터로 구현되며 다음과 같이 표현된다.

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

원 RASTA는 대역 필터로 작용하며, 고대역 부분은 주파수 영역에서의 장기적인 천이를 제거하여 CMS와 비슷한 역할을 한다. RASTA의 필터는 필터 계수 값에 따라 다양하게 적용할 수 있다. SRI Decipher에 사용된 RASTA 필터[4]는 CMS와 거의 비슷한 필터 특성을 보이며, 실시간 처리가 가능하다는 점 때문에 널리 사용되고, online CMS 버전의 일반화된 전달함수를 나타낸다.

$$H(z) = \frac{1 - z^{-1}}{1 - 0.97z^{-1}} \tag{5}$$

고대역 필터링은 파워 스펙트럼(power spectrum) 영역에 직접 적용하여 가산 잡음의 효과를 줄이는데 사용될 수 있다. Morgan과 Hermansky는 선형 필터와 가산 잡음을 결합하여 보상하는 방법으로 $J$ 값에 따라 선형 필터 또는 로그 필터의 특성을 갖는 linlog RASTA(또는 J-RASTA) 방법을 제안하였다[3]. 일반적으로 가산 잡음과 선형 필터의 효과(즉, 채널 잡음)를 동시에 완전하게 제거하지 못하기 때문에 두 잡음의 영향을 $J$ 값에 따라 실험적으로 결정하게 된다.

linlog RASTA 의 기본적인 알고리즘은 다음과 같다. (1)식에서 가산 잡음 $d(t)$가 영향을 준다면 음성 신호는

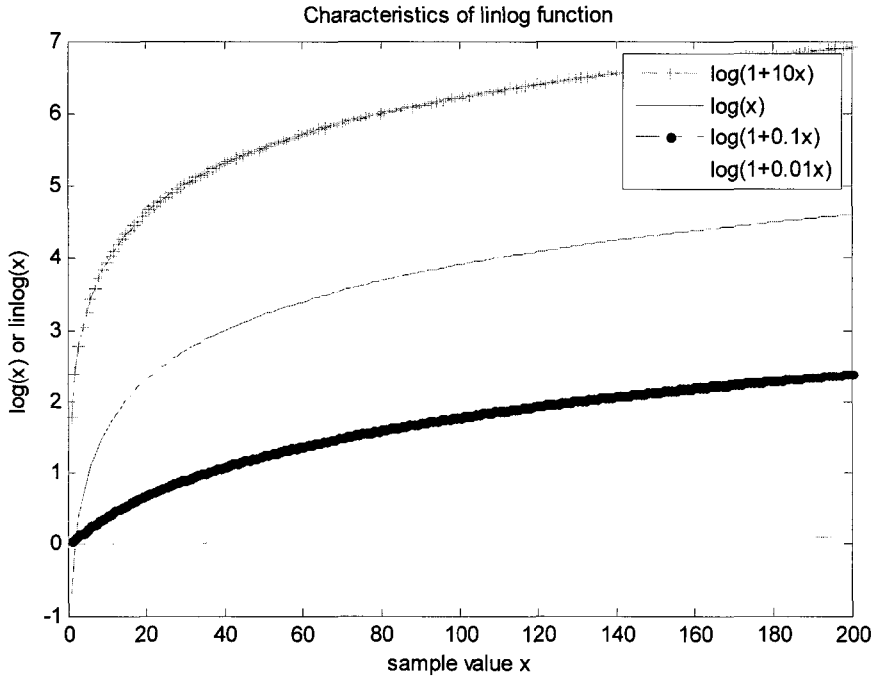
$$x(t) = f(t)^*(s(t) + d(t)) \tag{6}$$

로 표현될 수 있다. 파워 스펙트럼 영역으로 변환을 시키면 다음과 같이 되며, 간략화 시킬 수 있다.

$$\begin{aligned} X(w) &= F(w)(S(w) + D(w)) \\ &= F(w)D(w)(1 + \frac{1}{D(w)}S(w)) \\ &= \hat{F}(w)(1 + JS(w)) \\ &= \hat{F}(w)\hat{S}(w) \end{aligned} \tag{7}$$

위 식에서 보는 것과 같이 음성 $X(w)$는 선형 필터의 형식으로 모델링될 수 있기 때문에 가산 잡음에 종속적인 J값에 따라 채널 특성과 가산 잡음의 특성을 동

시에 모델링할 수 있다. 기존의 연구에서 가산 잡음이 거의 없는 음성에서는 $J=1e-6$인 경우에 가장 좋은 성능을 보였으며, 가산 잡음이 포함된 경우에는 SNR에 따라 $J$값을 결정하면(예; SNR 20dB인 경우 $J=10^{-7}$, SNR 10dB인 경우 $J=10^{-8}$, SNR 0dB인 경우 $J=10^{-9}$) 우수한 성능을 보인다고 발표하였다[3]. <그림 1>은 linlog 함수가 J값에 따라 선형, 또는 로그 특성을 나타내는 것을 보인다.



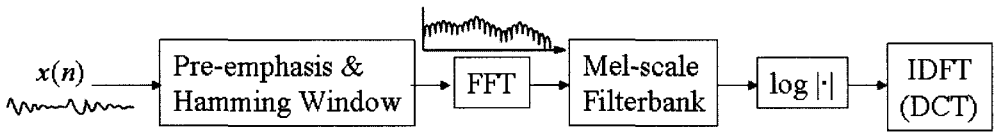

<그림 1> linlog 함수의 특성

위 그림에서 $J>>1$이면 log 함수의 특성을 보이며, $J<<1$이면 linear 함수의 특성을 보여, $J$값에 따라 선형 또는 로그 특성을 조절할 수 있다.

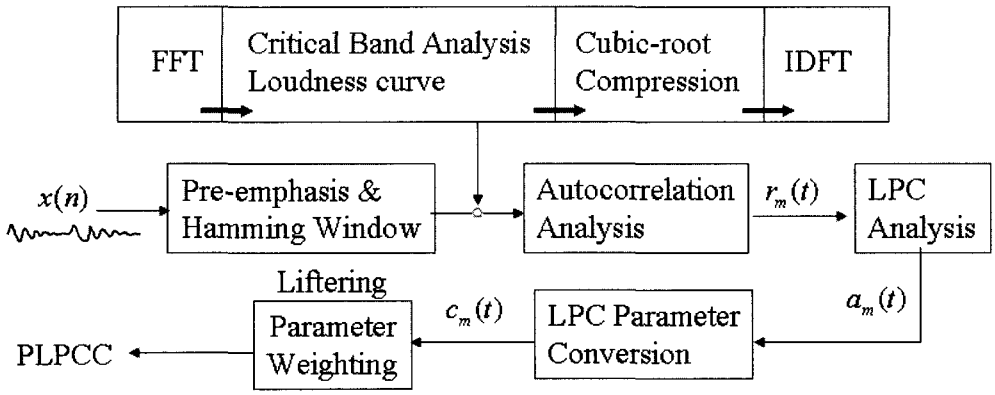
일반적으로 학습과 평가 시 동일한 $J$값을 사용해야 동일한 특성을 갖는 linlog 함수를 적용할 수 있다. 그러나 학습과 평가시 적용되는 가산 잡음의 양이 다를 경우, 동일한 $J$값을 이용하는 것이 성능의 저하를 가져올 수 있다. 따라서 학습시 동일한 데이터에 대해 여러 $J$값을 이용하여 특징 벡터를 만들어 학습 모델을 만든 다음, 평가시 특정 $J$값을 적용하는 방법을 사용하기도 한다[3].

## 3. Linlog MFCC

음성인식에 널리 사용되는 음성 특징 추출 방법으로는 MFCC와 PLP 방식이 있다. MFCC는 사람이 느끼는 피치의 단위인 멜(Mel)을 이용하여 주파수의 변화와 동일하게 느끼도록 주파수 축을 변환하여 음성 특징을 표현하는 방법으로 구현의 용이성과 높은 성능으로 많이 이용되고 있다. 반면에 PLP는 청취 과정의 심리음향학적 특성을 도입하여 모델링하는 방법으로 잡음 환경에서 성능이 우수하다고 알려져 있다. 잡음 환경에서 PLP 특징이 MFCC보다 성능이 우수하다고 하나, 소형 음성 인식 시스템이나 계산 능력이 부족한 환경에서는 MFCC를 이용한 잡음 환경에서의 음성 인식 시스템의 성능 향상이 필요하게 된다. <그림 2>는 MFCC와 PLP 특징의 추출 단계를 간략화하여 보여준다.




(a) MFCC 특징 표현

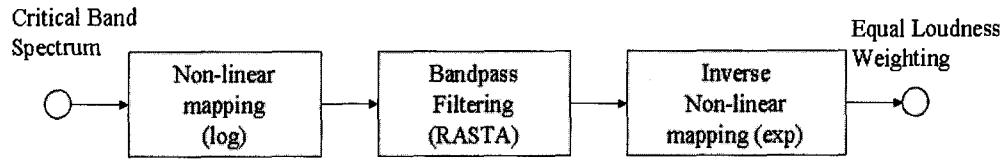



(b) PLP 특징 표현

<그림 2> MFCC와 PLP 특징의 표현 과정

위 그림에서 보는바와 같이 PLP 특징 추출은 MFCC 특징 추출 방법보다 많은 계산양과 단계를 필요로 한다. 기존의 발표된 RASTA-PLP는 PLP 특징 표현 단계에서 임계 밴드 분석(Critical Band Analysis)과정에서 비선형 함수를 적용한 후 대역 필터를 적용하고 다시 역변환 과정을 거친 후, 음량 조절(Equal loudness

weighting) 과정을 통과한다. 즉, 기존의 임계 밴드 스펙트럼에 대해 <그림 3>과 같은 과정을 추가로 적용한다.




<그림 3> PLP 특성에 RASTA 적용 과정

앞장에서 설명하였듯이 RASTA 특성은 CMS와 유사하게 입력 장치나 전달 경로의 선형적인 왜곡(선형 필터)를 제거하며 대역 필터의 역할을 한다. Linlog RASTA는 식 (7)에서와 같이 비선형함수를 $S(w)$가 아닌 $\hat{S}(w)$에 대해 적용함으로써 가산 잡음의 왜곡을 줄이도록 고안되었다. 대역 필터를 통과하여 선형 필터가 제거된 음성 신호의 파워 스펙트럼은 다음과 같이 표현된다.

$$Y(w) = \log(1 + JS(w)) \tag{8}$$

따라서 역 비선형 함수를 통과하여 다시 $S(w)$를 구하면

$$S(w) = \frac{e^{Y(w)} - 1}{J} \tag{9}$$

이 된다. 그러나 위 식은 음의 값을 가질 수 있기 때문에 기존의 linlog RASTA에서는 다음과 같은 근사치를 적용하여 모델링하였다.

$$\tilde{S}(w) = \frac{e^{Y(w)}}{J} \tag{10}$$

즉, $\tilde{S}(w) = S(w) + 1/J$의 관계를 갖는다.

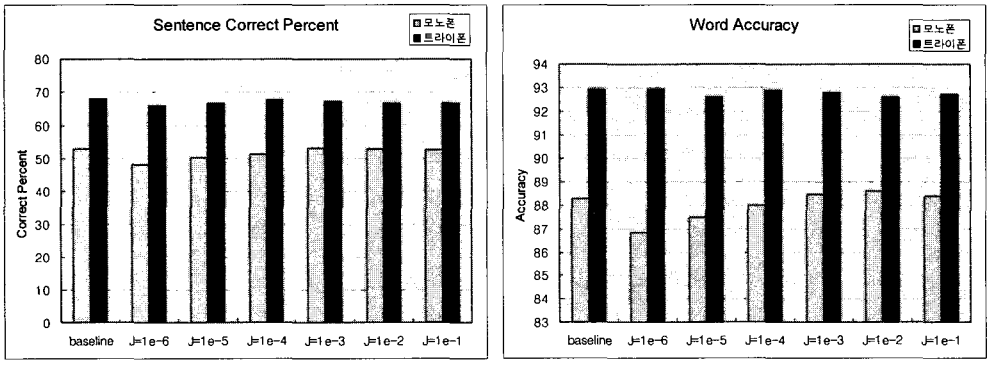
본 논문에서는 <그림 2>의 MFCC 특징 표현에서 log함수 대신 linlog RASTA에 적용된 식 (8)을 멜 스케일 필터뱅크의 크기(magnitude)와 파워 스펙트럼에 적용하였으며, 기존의 MFCC 특징 표현과 차이를 둬 제안된 방식을 linlog MFCC라고 하였다.

## 4. 실험 및 결과

기존의 연구에서 linlog RASTA 필터는 가산 잡음이 포함된 환경에서 *J*값에 따라 기존의 RASTA-PLP나 PLP에 비해 우수한 성능을 보였으며, 잡음이 포함되지 않은 음성에 대해서도 비슷한 성능을 보였다. 따라서, 본 논문에서는 가산 잡음이 포함되지 않고 채널 특성이 다른 환경과, 가산 잡음과 채널 잡음이 동시에 포함된 환경에서 인식 성능의 변화를 살펴 제안된 방식의 가능성을 파악하고자 한다.

첫 번째 실험은 학습 환경과 평가환경의 차이를 극대화하기 위하여 학습 자료는 조용한 환경에서 녹음된 DB를, 평가 자료는 전화 환경에서 녹음된 DB를 사용하였다. 훈련에 사용된 DB는 원광대에서 제작된 한국어 4연 숫자음성이며 지역적으로 고르게 분포된 400명이 발성하였고 16비트 16kHz로 샘플링되었다. 평가 음성 DB는 전화 환경에서 200명이 발성하였으며 유선/무선/Cellular/PCS 등의 4가지 분류로 8kHz 샘플링되었다. 학습 DB와 평가 DB의 환경을 유사하게 하기 위하여 훈련 DB를 8kHz로 다운 샘플링하여 실험하였다. 채널 잡음 보상 방법으로는 일반적으로 MFCC 특징을 이용하는 시스템에서 널리 사용하는 CMS를 이용하였다.

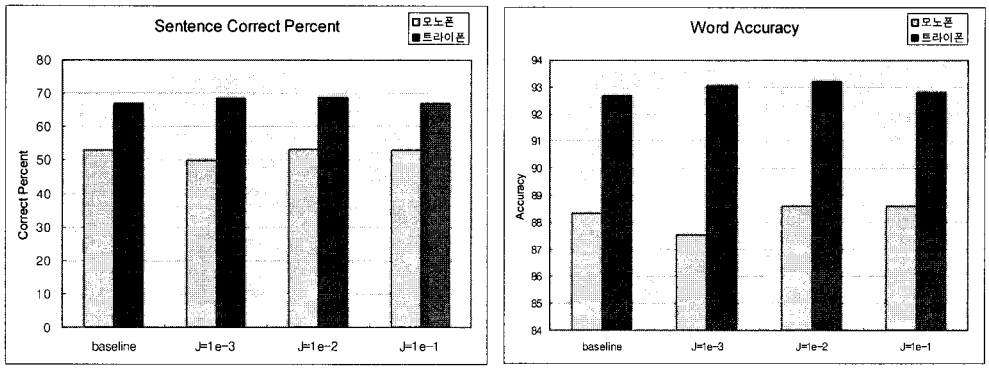
실험은 두가지로 나뉘어 실행되었다. 먼저 주파수 영역에서 파워스펙트럼을 이용하여 linlog 변환을 거친 후 켑스트럼 영역에서 CMS를 적용하여 파워스펙트럼에서의 linlog 함수의 특징을 살펴보았다. 다른 하나는 모든 과정은 동일하나 음성 신호의 파워스펙트럼이 아닌 크기에 linlog 함수를 적용하여 실험하였다. 실험 결과는 다음의 도표에서 정리하였다.




<그림 4> 음성 신호의 파워스펙트럼을 이용한 경우 J값에 따른 성능의 변화

실험 결과 음성 신호의 파워스펙트럼을 이용한 경우, 모노폰인 경우 *J=1e-3*에서 문장 인식률 성능이 향상되었으며, 단어 정확도의 경우 *J=1e-3, 1e-2, 1e-1*에서 성능

이 향상됨을 알 수 있다. 반면 트라이폰의 경우 다양한 $J$값의 변화에도 성능 향상은 이뤄지지 않았으며 기본 시스템과 거의 유사한 성능을 보였다.

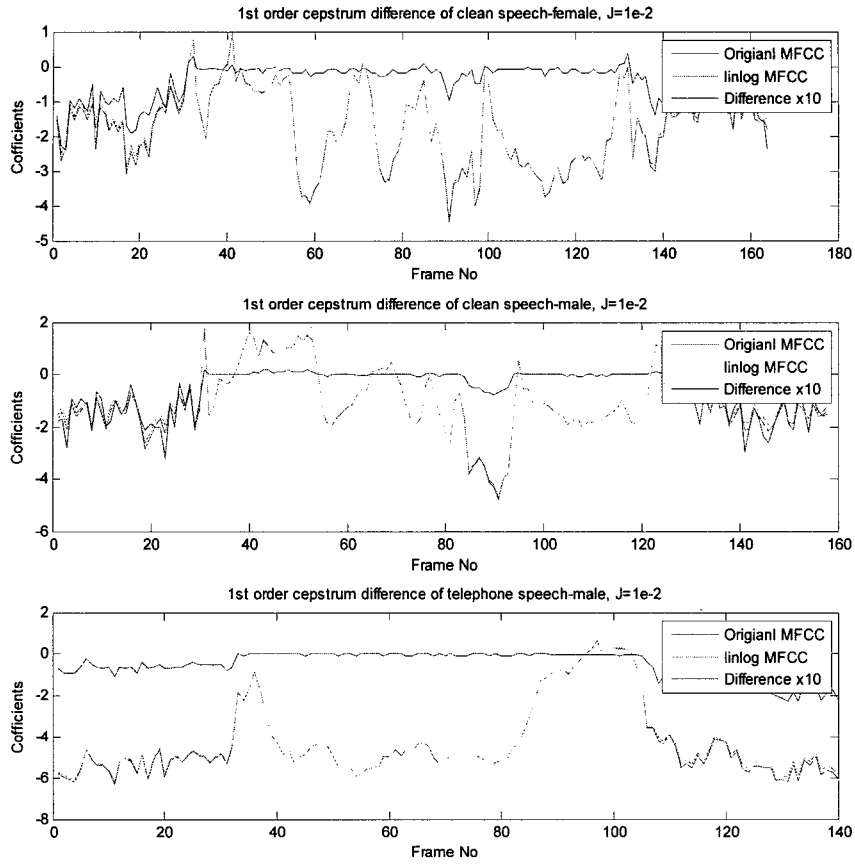



<그림 5> 음성 신호의 크기를 이용한 경우 J 값에 따른 성능의 변화

주파수 영역에서 음성 신호의 크기만을 이용한 경우에는 모노폰이나 트라이폰의 경우 모두 기본 시스템보다 성능이 향상됨을 알 수 있다. 특히 $J = 1e-2, 1e-1$ 값에서 문장 인식률이나 단어 정확도가 모두 향상되었다. Linlog 함수는 가산잡음과 채널잡음이 혼합된 경우에서 성능의 변화를 가져올 수 있다고 발표되었기 때문에 본 논문에서는 linlog MFCC의 적용가능성을 살펴보는데 주안점을 두었다. 위 실험결과에서 linlog 함수를 적용하여 조용한 환경에서 녹음된 음성 신호의 인식 성능이 저하되지 않고 오히려 향상됨으로 인하여 적용가능성을 볼 수 있다.

특히, <그림 4>에서 보고한 바와 같이 동일한 단어 정확도를 보이더라도 문장 인식률의 성능 차이가 발생하고 있다. 이는 발화된 음성 신호가 linlog 함수의 영향을 받아 음성 인식에 사용되는 특징 벡터의 변화를 가져오기 때문이라고 파악한다.

<그림 6>에서 보는 바와 같이 원 MFCC와 linlog MFCC의 차이를 확대하여 표시하면, 발화 음성의 시작 부분과 끝부분에서 차이가 발생함을 알 수 있다. 이는 발화 음성의 중간부분에서는 음성 신호의 크기가 커서 $J$값의 영향을 작게 받으나, 시작 부분과 끝부분에서는 주로 무성음이나 가산 잡음의 비율이 높기 때문에 영향의 폭이 커졌다고 생각할 수 있다.




<그림 6> 원 MFCC와 linlog MFCC의 차이

두 번째 실험은 ETSI의 AURORA2 DB를 이용하여, 채널 잡음과 가산 잡음이 함께 존재하는 경우에 제안된 방식의 성능 변화를 조사하였다. AURORA2 DB는 ETSI에서 DSR(distributed speech recognition) 시스템의 전단계(front-end) 알고리즘의 성능을 객관적으로 평가하기 위한 표준 데이터로 TI-DIGITS DB를 8kHz로 하향 샘플링하고 여러 가지 잡음과 선형 필터(채널 잡음) 효과를 주었다. 제공되는 DB 중 Set A와 B는 가산 잡음만을 고려한 것이며 Set C는 일반 전화 채널의 효과(MIRS 특성)를 필터링하고 "subway"와 "street"의 잡음을 첨가한 것이다. 부가된 잡음은 잡음 레벨(SNR 20, 15, 10, 5, 0, -5)에 따라 인위적으로 첨가되었다. Set C가 본 논문에서 제안하는 linlog MFCC의 목적에 적합하기 때문에 Set C를 이용한 실험을 수행하였다. 학습은 ETSI에서 권장하는 조용한 환경의 데이터를 사용하였으며(clean training), 첫 번째 실험과 동일한 조건을 부여하기 위하여 CMS를 적용하여 평가

실험을 하였다.

잡음 레벨과 linlog 함수의 상관 관계 특성을 파악하기 위하여 필터 뱅크의 출력을 크기와 파워 스펙트럼으로 구분하여 각각에 대해 $J$값을 변경하며 실험을 하였다. <표2>는 필터 뱅크의 출력이 크기인 경우의 인식 성능의 변화를 나타낸다.

<표 2> "Set c" 실험 결과: 특징 추출에 필터 뱅크 크기 값 이용(Accuracy 비교)

| system | place | clean | SNR 20 | SNR 15 | SNR 10 | SNR 5 | SNR 0 | SNR -5 |
|---|---|---|---|---|---|---|---|---|
| baseline | subway | 99.11 | 95.76 | 89.10 | 71.57 | 38.87 | 14.43 | 11.15 |
| | street | 99.15 | 96.34 | 90.84 | 73.37 | 44.35 | 18.38 | 9.76 |
| J=1e-2 | subway | 99.11 | 96.13 | 89.59 | 69.60 | 34.51 | 12.93 | 10.84 |
| | street | 99.27 | 96.52 | 90.08 | 70.71 | 43.17 | 17.59 | 8.49 |
| J=1e-3 | subway | 99.17 | 95.95 | 89.10 | 66.99 | 36.48 | 17.68 | 12.07 |
| | street | 99.12 | 95.89 | 88.66 | 69.47 | 42.84 | 20.53 | 10.07 |
| J=1e-4 | subway | 97.05 | 92.72 | 82.65 | 64.57 | 37.06 | 16.70 | 6.42 |
| | street | 97.07 | 92.47 | 86.16 | 68.08 | 42.32 | 20.41 | 6.17 |
| J=1e-5 | subway | 90.76 | 82.96 | 68.28 | 41.54 | 9.95 | -3.81 | -3.16 |
| | street | 91.14 | 85.70 | 76.45 | 56.08 | 31.29 | 12.03 | 3.72 |

위 실험 결과에서 $J$값이 1e-2, 1e-3인 경우에는 clean과 SNR 20에서 기본 시스템보다 성능이 향상됨을 알 수 있으며, $J$값이 1e-3, 1e-4인 경우 SNR 0과 -5에서 성능이 향상됨을 알 수 있다. 전체적인 경향을 살펴보면 $J = 1e-2, 1e-3$인 경우에 SNR이 높은 환경에서 성능이 향상되며, $J = 1e-3, 1e-4$인 경우에 SNR이 낮은 환경에서 성능이 향상됨을 알 수 있다. 그러나, $J$값이 더 작아지면 성능의 저하가 심해, 이 경우 잡음이 포함된 음성의 변별력을 급속히 떨어뜨리는 것으로 보인다.

동일한 실험을 필터 뱅크의 출력이 파워 스펙트럼인 경우에 진행하였다. 파워 스펙트럼을 이용한 경우, 잡음이 포함된 음성의 범위가 넓기 때문에, $J$ 값의 종류를 다양하게 하여 성능의 변화를 관찰하였다.

실험 결과, linlog MFCC는 기존의 linlog RASTA-PLP와 비슷하게 가산 잡음이 포함된 음성 신호의 채널 왜곡(선형 필터)를 제거하는데 사용될 수 있을 것으로 보이나, 기존의 linlog RASTA-PLP와는 조금 다른 경향을 보였다. 즉, 기존의 연구에서는 잡음의 양이 증가함에 따라 최고 성능을 보이는 $J$ 값도 작아지는 경향을 보였으나, 본 논문에서 제안하는 방식에서는 $J = 1e-6$인 경우에 각 환경에서의 성능이 향상됨을 알 수 있었다(평균 2.3%, ERR 7.1% 향상, 최고 7%, ERR 33% 향상). 이는 다음과 같이 해석할 수 있다. 첫째는 기존 연구에서는 주파수 영역에서 비선형 함수(linlog)를 적용한 후, RASTA 필터를 통과하였으나, 본 논문에서는 비선형 함수를 통과한 후 DCT 과정을 거친 켑스트럼 벡터에 대해 CMS를 적용하였기 때문으로 해석할 수 있으며, 대역 필터의 특성을 갖는 RASTA와 고대역 필터 역할하는 CMS의 특성 차이가 영향을 줄 수도 있다. 두 번째로는 기존의 연구에서 linlog RASTA 필터는 주파수 영역의 임계 대역(critical band) power spectrum에 대해

적용되었으나, 본 논문에서는 linlog 함수가 멜 필터뱅크의 출력에 적용되어 경향이 달라질 수 있다. 이런 해석은 추후 연구를 통해 명확하게 검증되어야 할 것이다.

<표 3> "Set c" 실험 결과: 특징 추출에 필터 뱅크 파워 스펙트럼 이용
(Accuracy 비교)

| system | place | clean | SNR 20 | SNR 15 | SNR 10 | SNR 5 | SNR 0 | SNR -5 |
|---|---|---|---|---|---|---|---|---|
| baseline | subway | 99.14 | 95.89 | 89.38 | 72.15 | 39.15 | 14.58 | 11.42 |
| | street | 99.15 | 96.34 | 91.14 | 74.09 | 44.92 | 18.29 | 9.89 |
| J=1e-2 | subway | 99.14 | 95.12 | 86.89 | 65.27 | 33.22 | 14.74 | 11.39 |
| | street | 99.12 | 96.19 | 89.48 | 70.34 | 40.45 | 17.62 | 10.64 |
| J=1e-3 | subway | 99.11 | 94.96 | 87.35 | 64.6 | 32.76 | 13.66 | 10.62 |
| | street | 99.21 | 96.01 | 89.33 | 70.04 | 42.11 | 18.08 | 9.82 |
| J=1e-4 | subway | 99.11 | 94.26 | 85.94 | 64.35 | 31.53 | 13.23 | 11.45 |
| | street | 99.12 | 95.01 | 87.09 | 66.75 | 40.54 | 17.87 | 10.07 |
| J=1e-5 | subway | 99.26 | 95.49 | 89.13 | 67.39 | 34.33 | 13.54 | 11.64 |
| | street | 99.12 | 96.25 | 89.66 | 70.31 | 42.59 | 17.35 | 9.58 |
| J=1e-6 | subway | 99.17 | 96.62 | 92.88 | 78.32 | 46.21 | 19.13 | 11.21 |
| | street | 99 | 96.43 | 91.6 | 75.03 | 48.25 | 22.73 | 10.58 |
| J=1e-7 | subway | 97.57 | 93.68 | 86.06 | 69.91 | 43.17 | 19.93 | 10.65 |
| | street | 97.73 | 92.99 | 85.13 | 68.44 | 43.68 | 20.89 | 8.68 |
| J=1e-8 | subway | 90.88 | 80.87 | 65.77 | 43.32 | 19.8 | 6.08 | -2.33 |
| | street | 91.14 | 84.76 | 74.73 | 54.78 | 33.04 | 13.81 | 1.93 |
| J=1e-9 | subway | 78.91 | 68.96 | 52.16 | 21 | -1.72 | -8.38 | -4.61 |
| | street | 78.6 | 73.4 | 63.33 | 45.89 | 21.55 | 8.16 | 2.84 |

## 5. 결 론

본 논문에서는 채널 잡음과 가산 잡음이 포함된 음성의 신호를 비선형 함수를 통하여 보상하는 linlog RASTA 연구를 MFCC 특징에 적용하여 그 가능성을 살펴보았다. RASTA-PLP나 linlog RASTA-PLP 모두 선형필터에 의한 채널 왜곡을 보상하는데 사용되나, PLP 특징에 기반하고 있기 때문에 특징 추출 과정에서 많은 연산량과 단계가 필요하다. 그러나 본 논문에서 제안하는 linlog MFCC는 계산량이 기존의 MFCC와 거의 동일하며, 그 경향은 기존의 linlog RASTA와 유사하기 때문에 채널과 가산 잡음이 포함된 환경에 적용할 수 있을 것이다. 특히 환경에 적합한 *J*값을 미리 선정할 수 있다면 포함된 잡음의 성격을 미리 파악하지 않고 채널 잡음과 가산 잡음을 어느 정도 보상할 수 있을 것으로 본다. 따라서 추후 연구로는 실험과정에서 생긴 의문점을 해소하기 위하여 켑스트럼 영역에서 CMS를 적용하는 대신, 주파수 영역에서 RASTA 필터를 적용하여 성능 변화를 확인하고, 또한 J 값이 채널 잡음과 가산 잡음을 동시에 보상하는데 중요한 역할을 하기 때문에 환경에 적응하는 *J* 값을 예측하거나, 잡음의 특성을 파악하여 적절한 *J* 값을 결정하는 연구를 진행할 예정이다.

# 참 고 문 헌

[1] H. Hermansky, "Perceptual linear predictive(PLP) analysis of speech", *Journal of the Acoustical Society of America*, vol. 87, no. 4, pp. 1738-1752, Apr. 1990.

[2] H. Hermansy, N. Morgan, et al., "RASTA-PLP speech analysis", TR-91-069, International Computer Science Institute, Berkley, Dec. 1991.

[3] H. Hermansky, N. Morgan and H.-G. Hirsch, "Recognition of speech additive convolutional noise based on RASTA spectral processing", *Proc. ICASSP*, vol. 2, pp. 83-86, 1993.

[4] F. H. Liu, R. M. Stern, et al., "Efficient cepstral normalization for robust speech recognition", *Proc. ARPA Speech and Nat. Language Workshop*, Princeton, NJ, pp. 69-74, Mar. 1993.

[5] H. Ogawa, "More robust J-RASTA processing using spectral subtraction and harmonic sieving", TR-97-031, International Computer Science Institute, Berkley, Aug. 1997.

접수일자: 2006년 8월 1일
게재결정: 2006년 9월 23일

▶ 윤영선(Young-Sun Yun)
주소: 306-791 대전광역시 대덕구 오정동 133번지 한남대학교
소속: 한남대학교 정보통신공학과
전화: 042) 629-7569
E-mail: ysyun@hannam.ac.kr