

캡스트럼으로부터 변환된 로그 스펙트럼을 이용한 포먼트 평활화 캡스트럴 평균 차감법

Formant-broadened CMS Using the Log-spectrum Transformed from the Cepstrum

정혜경*, 김유진**, 정재호**
(Hea-Kyoung Jung*, Yu-Jin Kim**, Jae-Ho Chung**)

*국방과학연구소 4체계 개발본부 전자전체계부 2팀, **인하대학교 전자공학과 디지털 신호처리 연구실
(접수일자: 2001년 11월 12일; 채택일자: 2002년 1월 28일)

본 논문에서는 음성 인식과 화자 인식에서 채널 변이 정규화를 위해 널리 사용되는 전통적인 캡스트럴 평균 차감법 (CMS; Cepstral Mean Subtraction)의 성능을 향상시키기 위한 정규화 방법을 제안한다. 기존의 캡스트럴 평균 차감법은 장구간 캡스트럼의 평균으로 채널 성분을 추정하므로 유성음의 포먼트에 의해 채널 성분이 편향되는 단점을 가진다. 제안된 포먼트 평활화 캡스트럴 평균 차감법 (FBCMS; Formant-broadened CMS)은 캡스트럼으로부터 변환된 로그 스펙트럼에서 포먼트 위치를 쉽게 찾을 수 있고, 포먼트는 전극점 모델로 표현되는 성도 전달 함수의 우세 극점에 대응된다는 사실에 근거한다. 따라서 제안된 방법은 캡스트럼으로부터 음성의 포먼트를 구하고, 이로부터 포먼트의 대역폭을 확장한 캡스트럼을 구한 후 평균함으로써 채널 캡스트럼 성분으로부터 우세 극점들의 영향을 제거한다. 전극점 모델의 우세 극점을 얻기 위해 다항식 인수분해 과정을 거치지 않으므로 연산량을 줄일 수 있으며 포먼트에 해당하는 우세 극점만으로 선택적으로 처리할 수 있다. 본 연구에서는 4가지의 모의 채널을 이용하여 전통적인 캡스트럴 평균 차감법, 극점 필터화 캡스트럴 평균 차감법 (Pole-filtered CMS) 그리고 제안된 방법의 비교실험을 수행하였다. 실제 채널 캡스트럼과 추정된 채널 캡스트럼과의 거리를 측정하는 실험에서 음성에 의한 편향을 완화시켜 실제 채널에 보다 가까운 평균 캡스트럼을 얻을 수 있음을 확인하였다. 또한 문장 독립 화자 식별에서 제안된 방법은 전통적인 캡스트럴 평균 차감법보다 우세하고 극점 필터화 캡스트럴 평균 차감법 (Pole-filtered CMS)과는 비슷한 결과를 보였다. 결과적으로 제안된 방법은 전통적인 캡스트럴 평균 차감법에 기반하여 효과적인 채널 정규화가 가능하다는 것을 보였다.

핵심용어: 화자인식, 음성인식, 채널잡음제거, 채널불일치 정규화, 캡스트럴 정규화, 캡스트럴 평균 차감법
투고분야: 음성처리 분야 (2.5)

In this paper, we propose a channel normalization method to improve the performance of CMS (cepstral mean subtraction) which is widely adopted to normalize a channel variation for speech and speaker recognition. CMS which estimates the channel effects by averaging long-term cepstrum has a weak point that the estimated channel is biased by the formants of voiced speech which include a useful speech information. The proposed Formant-broadened Cepstral Mean Subtraction (FBCMS) is based on the facts that the formants can be found easily in log spectrum which is transformed from the cepstrum by fourier transform and the formants correspond to the dominant poles of all-pole model which is usually modeled vocal tract. The FBCMS evaluates only poles to be broadened from the log spectrum without polynomial factorization and makes a formant-broadened cepstrum by broadening the bandwidths of formant poles.

We can estimate the channel cepstrum effectively by averaging formant-broadened cepstral coefficients. We performed the experiments to compare FBCMS with CMS, PFCMS using 4 simulated telephone channels. In the experiment of channel estimation, we evaluated the distance cepstrum of real channel from the cepstrum of estimated channel and found that we were able to get the mean cepstrum closer to the channel cepstrum due to an softening the bias of mean cepstrum to speech. In the experiment of text-independent speaker identification, we showed the result that the proposed method was superior than the conventional CMS and comparable to the pole-filtered CMS. Consequently, we showed the proposed method was efficiently able to normalize the channel variation based on the conventional CMS.

Keywords: Speaker recognition, Speech recognition, Channel noise removal, Channel mismatch normalization, Cepstral normalization, CMS

ASK subject classification: Speech signal processing (2,5)

I. 서론

다양한 채널 환경에 강인한 음성 인식 및 화자 인식에 관한 연구는 실험실 환경에서 연구되어온 인식 시스템을 실용화하기 위해 진행되고 있으며 인식 성능을 개선시킬 수 있는 여러 가지 방법들이 제안되어 왔다. 이는 음성 인식이나 화자 인식 시스템이 훈련과 테스트 환경이 일치하지 않는 불일치 조건에서는 일치된 환경에 비하여 크게 성능이 저하되기 때문이다[1-3].

켄스트럴 평균 정규화 (CMN; Cepstral Mean Normalization) 또는 켄스트럴 평균 차감법 (CMS; Cepstral Mean Subtraction)은 전화선 환경에서 음성 인식이나 화자인식 시스템의 성능을 향상시키기 위해 사용되어온 방법이며 훈련환경과 테스트 환경의 불일치를 켄스트럴 영역에서 정규화하는 채널 정규화 기법으로 널리 알려져 있다. CMS는 음향학적인 분포가 고른 음성의 장구간 평균 켄스트럴은 상쇄되어 제거되고 시간에 대해 일정한 채널 성분만이 남는다는 가정하에 채널을 통과한 켄스트럴의 장구간 평균을 차감함으로써 이루어진다. 단순한 방법에도 불구하고 전통적인 CMS는 훈련과 테스트 환경에 모두 적용되었을 때 큰 인식 성능의 향상을 나타내며 특히 훈련과 테스트에서 동일한 발성을 사용한 경우 큰 효과를 거둘 수 있다. 기본적인 CMS

의 구조는 그림 1과 같이 켄스트럴 변환, 채널 추정 그리고 추정된 채널 성분의 차감 과정으로 구성된다.

그러나 CMS에서의 채널추정 과정에서는 실제로 발생되는 음성의 음향학적 분포가 고르지 않다는 점을 고려하지 않고 있다. 따라서 켄스트럴법의 장구간 평균에서 음성 성분은 모두 상쇄되지 않고 남게 되어 채널 성분과 함께 차감되므로 화자 및 음성 인식에 유용한 정보가 손실되는 결과를 초래한다. 이러한 단점을 극복하기 위해 최근까지 보다 효과적인 채널 성분을 추정하기 위한 연구가 진행되어 왔다. MMCMNFW[4]는 로그 스펙트럼으로부터 채널 켄스트럴을 추정하는 방법을, PFCMS[5,6]는 선형 예측 모델에서의 극점을 필터링하는 방법을, CSAR[7]은 훈련과 테스트의 발성이 유사하고 스테레오 데이터를 사용할 수 있을 경우에 복소 켄스트럴을 통해 채널 성분을 추정하는 방법을, 그리고 DPCMS[8]에서는 채널 특성에 민감한 특정 대역의 극점을 제거하여 채널 성분을 추정하는 방법을 각각 제안하였다.

한편 이러한 연구 가운데 Naik의 PFCMS는 전통적인 CMS의 간단한 기본 구조를 그대로 유지하면서 스테레오 데이터와 같은 사전 정보가 없이 채널 성분을 효과적으로 추정할 수 있는 방법으로 알려져 있다. 그러나 PFCMS의 기본 이론은 협대역 극점을 찾기 위해 선형 예측 다항식

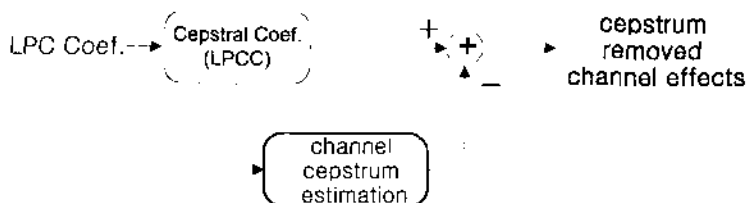


그림 1. 켄스트럴 평균 차감법의 기본 구조
Fig. 1. The basic framework of Cepstral Mean Subtraction.

으로부터 모든 극점을 구하는 많은 연산량을 요구하는 단점을 가지고 있다.

본 논문에서는 채널정규화를 위해 널리 사용되는 CMS의 성능을 향상시키기 위해 PFCMS에서 제안된 극점 필터링의 결과를 얻을 수 있고 연산량 관점에서 좀더 효과적인 포먼트 대역확장 CMS (FBCMS; formant-broadened CMS)에 대해서 설명한다.

2장에서는 채널 정규화 기법의 이론적인 바탕인 호모몰픽 디컨벌루션 (Homomorphic Deconvolution)에 대해 설명하고, 전통적인 CMS 알고리즘과 이를 향상시키기 위해 제안된 PFCMS에 대해 간단히 살펴본다. 3장에서는 제안된 포먼트 대역 확장 CMS를 설명하고 4장에서는 제안된 방법과 기존의 방법을 비교하는 실험에 대해 기술하고 그 결과를 고찰한다. 마지막으로 5장에서 결론과 앞으로의 연구방향을 제시한다.

II. 채널 정규화 기법

2.1. 호모몰픽 디컨벌루션

일반적으로 음성 신호는 호모몰픽 시스템의 관점에서 살펴보면 음향학적인 변이에 따라 빠르게 변하는 성도 신호와 이에 반해 상대적으로 거의 변하지 않거나 매우 느리게 변하는 여기 신호로 구성되어 있다. 호모몰픽 디컨벌루션은 음성의 이러한 특성을 이용하여 여기 신호와 성도 신호를 분리해 내는 분석방법이며 호모몰픽 필터링이라고 하는 비선형적인 필터링 방법으로도 알려져 있다. 컨벌루션된 신호를 분리하고 복원하기 위한 일반적인 호모몰픽 시스템을 그림 2에 나타내었다[9]. 그림에서 첫 번째 시스템인 D_* 는 입력 신호 $x[n]$ 을 출력신호 $\hat{x}[n]$ 으로 다음과 같은 복소 로그리즘 관계에 의해 대응시킨다.

$$\hat{X}(z) = \log X(z) \tag{1}$$

이때 복소 로그리즘은 $\log X(z) = \log|X(z)| + j\arg[X(z)]$ 으로 정의되고, $X(z)$ 와 $\hat{X}(z)$ 는 각각 $x[n]$ 과 $\hat{x}[n]$ 의 z

변환이다. 만약 $x[n]$ 이 다음과 같은 신호 $s[n]$ 과 채널의 시스템 함수 $h[n]$, 두 신호의 컨벌루션이라면 다음과 같이 표현될 수 있다.

$$x[n] = s[n] * h[n] \tag{2}$$

이때 $*$ 는 두 신호의 이산 컨벌루션이며 따라서

$$\begin{aligned} \hat{X}(z) &= \log(X(z)) \\ &= \log(S(z) \cdot H(z)) \\ &= \log(S(z)) + \log(H(z)) \\ &= \hat{S}(z) + \hat{H}(z) \end{aligned} \tag{3}$$

또는 시간 축에서 다음과 같이 표현될 수 있다.

$$\hat{x}[n] = \hat{s}[n] + \hat{h}[n] \tag{4}$$

따라서 시스템 D_* 는 컨벌루션된 두 신호를 합 의 형태로 변환시키게 된다. 이러한 첫 번째 시스템의 결과를 복소 켈스트럼이라고 부른다. 이때 $s[n]$ 을 음성의 여기 신호, $h[n]$ 을 음성의 시스템 함수 즉 성도의 단구간 채널 함수라고 가정하면 가산 형태의 두 켈스트럼 신호는 각각 다른 켈스트럼 영역, 즉 큐프렌시 (quefrensy) 영역에 분포하게 된다. 따라서 가중 함수를 이용한 리프터링 (liftering)을 통해서 쉽게 성도 채널 함수만을 추출할 수 있고 일반적으로 추출된 켈스트럼 신호는 단구간 음성을 모델링하기 위한 음성 켈스트럼으로 사용하고 있다. 만약 가산 형태의 두 켈스트럼 신호가 동일 큐프렌시 영역에 존재할 경우 차감 연산을 통해서 잡음을 최소화할 수 있다. 이러한 특징은 로그 스펙트럼 영역에서의 필터링 또는 켈스트럼 영역에서의 리프터링을 통해서 채널의 특징을 제거하거나 보상할 수 있다는 것을 의미한다[10-12].

두 번째 시스템인 D_*^{-1} 시스템은 호모몰픽 시스템에 의해 변환된 복소 켈스트럼 $\hat{x}[n]$ 을 원하는 선형 필터 연산을 거쳐 만들어진 $\hat{y}[n]$ 을 다시 컨벌루션 형태의 신호로 복원하게 된다. 두 번째 시스템에 의한 처리과정은 다음과 같이 나타낼 수 있다.

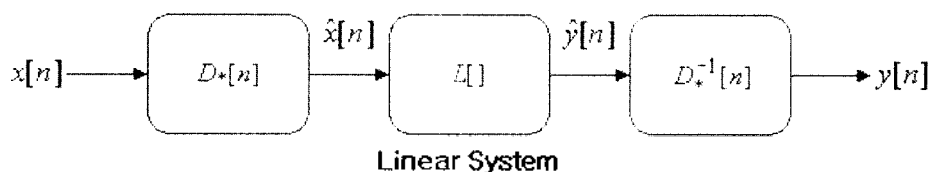


그림 2. 디컨벌루션을 위한 호모몰픽 시스템
Fig. 2. The system for homomorphic deconvolution.

$$D_s^{-1}(\hat{y}[n]) = y[n] \quad (5)$$

$$Y(z) = \exp(\hat{Y}(z))$$

이때 $Y(z)$ 와 $\hat{Y}(z)$ 는 각각 $y[n]$ 과 $\hat{y}[n]$ 의 z 변환이며 D_s^{-1} 시스템은 D_s 시스템과 함께 사용되어 보코더 등의 시스템을 구성할 수 있으며 음성 인식 또는 화자 인식 등의 특징을 처리하는 과정에서는 생략될 수 있다. 결론적으로 호모몰픽 디컨벌루션은 컨벌루션 형태로 결합된 두 개 이상의 신호에 대해서 각각의 신호로 분리해 내거나 특정 신호를 위한 처리를 가능토록 한다.

2.2. 켈스트럼 평균 차감법

CMS는 CMN (cepstral mean normalization)이라고도 하며 시간 영역에서 컨벌루션의 성격을 띠는 채널 잡음이 켈스트럼 영역에서는 가산적인 형태가 된다는 점을 이용하여, 켈스트럼 영역에서 채널 효과를 정규화하는 것이 목적이다. CMS의 기본 이론은 깨끗한 음성의 장구간 켈스트럼 평균, $\hat{x}_{avg}[n]$ 이 0이 된다는 가정하에 채널통과 음성의 켈스트럼 평균을 구하면 채널의 켈스트럼만 남는다는 이론을 바탕으로 한다.

채널을 통과한 음성의 장구간 평균은 다음과 같이 표현될 수 있다.

$$\hat{x}_{avg}[n] \cong \frac{1}{N} \sum_{i=1}^N \hat{x}_i[n]$$

$$= \frac{1}{N} \sum_{i=1}^N \hat{s}_i[n] + \frac{1}{N} \sum_{i=1}^N \hat{h}_i[n] \quad (6)$$

$$\cong \hat{s}_{avg}[n] + \hat{h}_{avg}[n]$$

이다. 이때 N 은 음성의 프레임 수이다. 일반적으로 발생 구간 내에서의 채널 특성은 급격하게 변화하지 않는다면 $\frac{1}{N} \sum_{i=1}^N \hat{h}_i[n]$ 은 시간 순서와 무관한 $\hat{h}[n]$ 으로 표현할 수 있다. 또한 깨끗한 음성이 장구간에서 균일한 분포(유무성음의 비율, 발생 길이 등의 부분에서)를 보인다면 켈스트럼은 서로 상쇄되어 평균벡터가 0에 가깝다고 할 수 있다. 따라서 $\hat{s}_{avg}[n] \Rightarrow 0$ 이라고 가정한다면 다음과 같이 표현할 수 있으며

$$\hat{x}_{avg}[n] = \hat{h}_{avg}[n] + \hat{s}_{avg}[n]$$

$$= \hat{h}_{avg}[n] + 0 \quad (7)$$

$$= \hat{h}[n]$$

이것은 채널만의 켈스트럼 성분이 된다. 따라서 채널통과 음성의 켈스트럼 평균, 즉 채널의 켈스트럼을 구하여 채널통과 음성으로부터 제거하면 채널의 영향을 받지 않은 깨끗한 음성의 켈스트럼을 다음과 같이 구해낼 수 있다.

$$\hat{x}_{cms,i}[n] = \hat{x}_i[n] - \hat{h}_{avg}[n]$$

$$= (\hat{s}_i[n] + \hat{h}[n]) - \hat{h}[n] \quad (8)$$

$$= \hat{s}_i[n]$$

즉 CMS의 결과 벡터인 $c_{cms,i}$ 는 채널 켈스트럼에 영향을 받지 않는 깨끗한 신호라 할 수 있다.

그러나 실제상황에서 사용할 수 있는 훈련 데이터와 테스트 데이터에는 음소의 양과 종류에 있어서 한계가 있으므로 음성의 장구간 켈스트럼 평균이 0이 된다는 가정은 적용되기가 힘들다. 유, 무성음과 마찰음, 파열음 등이 음성 전체 구간 내에서 고르게 분포할 경우 켈스트럼 계수들의 평균값은 0에 가깝지만 보통의 발성의 경우, 특히 한국어와 같은 발성에서는 고른 분포를 가지기란 쉽지 않다. 따라서 CMS에서 추정된 채널은 유용하게 사용될 수 있는 음성정보까지 포함하게 되므로 채널 영향을 제거하는 것은 화자인식이나 음성인식의 정확도를 떨어뜨리는 요인으로 작용한다. 그러므로 적절한 채널 정규화를 위해서는 채널을 거친 음성으로부터 채널 추정시에 음성정보를 효과적으로 분리하는 것이 필요하다.

2.3. 극점 필터화 켈스트럼 평균 차감법 (PFCMS: Pole-Filtered CMS)

채널 정규화를 위한 연구들 가운데 PFCMS는 기존 CMS의 구조를 기반으로 채널에 대한 사전정보 또는 스테레오 데이터가 없이도 채널을 추정할 수 있는 효과적인 방법으로 알려져 있다. 전극점 모델에서 우세한 현대역 극점은 주로 음성 구간에서 나타나고 평균에 의한 채널 추정에서 채널 켈스트럼을 편향시킨다. PFCMS 방법은 선형예측 모델로부터 현대역 극점을 필터링을 함으로써 기존의 CMS 방법에서 음성에 의한 평균 켈스트럼의 편향을 최소화시킨다. 즉 PFCMS는 우세 극점들의 영향을 줄인 후 켈스트럼 평균을 구함으로써 채널 추정시에 잔여 음성에 의한 편향을 최소화하고 장구간 음성 성분 평균이 0에 가깝도록 하여 켈스트럼 차감법의 효과를 높이고자 한다.

PFCMS는 p 차의 근 $z_k, k=1, 2, 3, \dots, p$, 를 가지는 선형 예측 분석에 의한 전극점 모델에 기반하며 이 모델에서 근들은 음성을 모델링하는데 있어서 우세한 부분을 형성한다.

$$S(z) = \frac{1}{A(z)} = \frac{1}{1 + \sum_{k=1}^p a_k z^{-k}} \quad (9)$$

$$= \frac{1}{\prod_{k=1}^p (1 - z_k z^{-1})}$$

Naik는 단위 원에 가까우면서 협대역을 가진 선형 예측 극점은 채널이나 잡음의 영향에 덜 민감하고 유용한 음성의 정보를 많이 포함하고 있으면서 채널 성분 추정에 큰 영향을 미친다는 것을 보여주었다. 따라서 이런 협대역 극점의 대역을 넓히기 위해 선형예측 분석에서 극점을 수정하는 작업, 즉 극점 필터링을 통해서 중심 주파수를 유지하면서 단위 원으로부터 원점에 가까워지도록 근을 이동하는 일을 수행하였다[4-6].

그림 3은 반경 1인 단위원에서 극점을 문턱값 처리(문턱값으로 강제 보정)하는 과정을 보여준다. 선형 예측 다항식에서 구할 수 있는 근 중 그 절대값이 일정 문턱값(점선) 이상인 근을 우세극점이라고 하고 그 크기를 문턱값 이하로 내려주는 것이다. 이때 우세극점의 주파수는 그대로 유지하면서 크기만 이동함으로써 극점 필터링을 한다. 이런 과정을 거쳐 협대역 극점의 영향을 제거한 후 얻은 켄스트럼 계수를 극점 필터링 켄스트럼 계수(PFCC: pole-filtered cepstral coefficient)라 하고, PFCC의 평균으로 얻을 수 있는 켄스트럼 계수를 추정된 채널로 정의하여 CMS에서 사용하는 켄스트럼 평균 대신에 차감해 줌으로써 채널을 통과하지 않은 음성의 켄스트럼 계수에 좀더 가까운 특징 파라미터를 얻어낼 수 있다.

PFCCMS의 구체적인 과정은 그림 4(a)에서 설명할 수 있

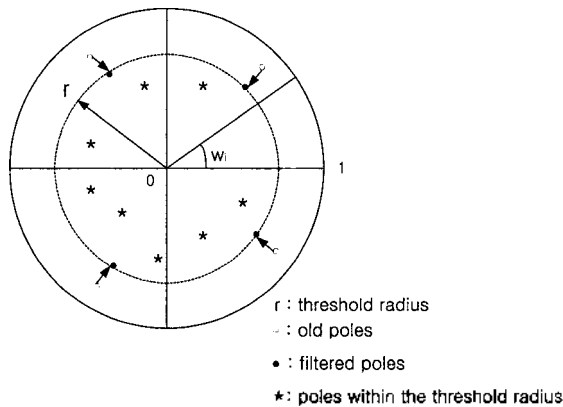


그림 3. 단위원에서의 극점 문턱값 처리
Fig. 3. Pole thresholding process on the unit circle.

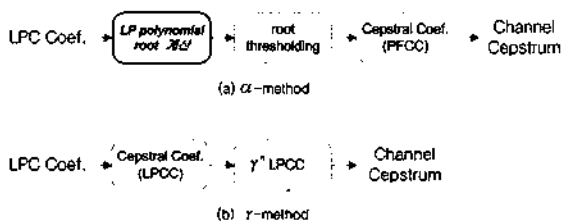


그림 4. PFCMS에서의 채널추정 순서도
Fig. 4. Channel estimation in PFCMS.

다. 즉, 선형예측 계수로 이루어진 다항식에서 직접적으로 근을 구하여 근의 주파수는 변화시키지 않으면서 크기만을 α 만큼 문턱값 처리함으로써 극점을 필터링하는 것으로 이 방법을 α -방법이라 한다. α -방법에서 근을 구하는 과정의 과다한 연산을 피하기 위하여 극점의 대역을 넓힐 수 있는 또다른 방법으로써 선형 예측 켄스트럼 계수(LPCC)를 직접적으로 조정하여 극점 필터링의 효과를 얻는 방법이 있다. 그림 4(b)에서 볼 수 있는 것처럼 켄스트럼 계수에 γ^n 을 곱하여 PFCC를 얻는 것으로서, γ -방법이라 한다. 이때 문턱값을 넓혀줄 대역폭 δ 값을 고려하여 $\gamma = e^{-\frac{\delta}{2}}$ 에 의해서 결정될 수 있다. γ -방법은 α -방법에 비해 연산량이 적은 장점을 가진 반면, 극점 필터링의 효과를 γ 값의 조정에만 의존해야 하므로 특정 극점에 대해서 원하는 만큼의 정확한 필터링이 어려운 단점이 있다.

최근에 α -방법과 γ -방법의 연산량과 채널 추정 성능을 비교한 연구결과가 있었으며 화자확인 시스템에 적용했을 경우에 α -방법의 과다한 연산을 지적하고 γ -방법에서의 연산과 성능향상을 위한 효과적인 극점 필터링의 문턱값에 대한 연구결과가 발표되었다[13].

III. 포먼트 대역 확장 켄스트럼 평균 정규화

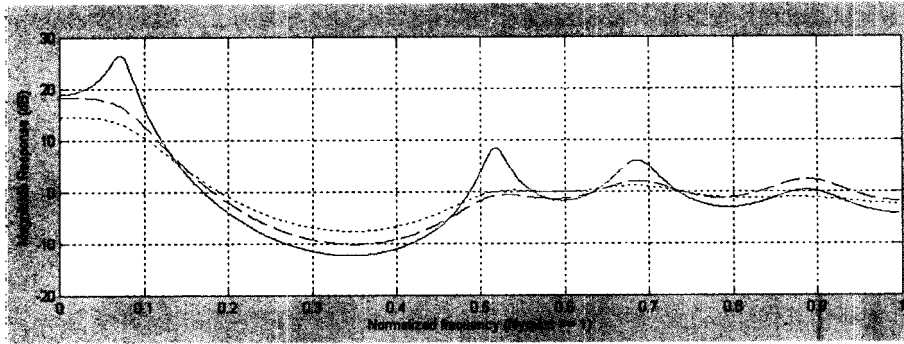
3.1. 우세 극점 필터링과 협대역 포먼트 대역 확장

선형 예측 시스템의 우세 극점 필터링 결과는 단 구간에서 선형 예측 시스템인 성도 모델의 주파수 응답에서 확인될 수 있다. 이를 확인하기 위해 음성의 모음 부분에 대한 α -방법과 γ -방법에 의한 극점 필터링 처리를 수행하였다. 각 방법들을 위해 0.85의 문턱값과 선형 예측 분석을 통해 얻어진 12차 예측 계수를 사용하였고 결과의 주파수 응답을 그림 5(a)에서 나타내었다. 그림을 통해 필터링된 우세 극점들의 영향이 주로 음성의 포먼트 대역을 평활화하는 결과로 나타나는 것을 확인할 수 있다. 이러한 현상은 전달함수의 극점이 음성의 포먼트에 대응되기 때문이다[10].

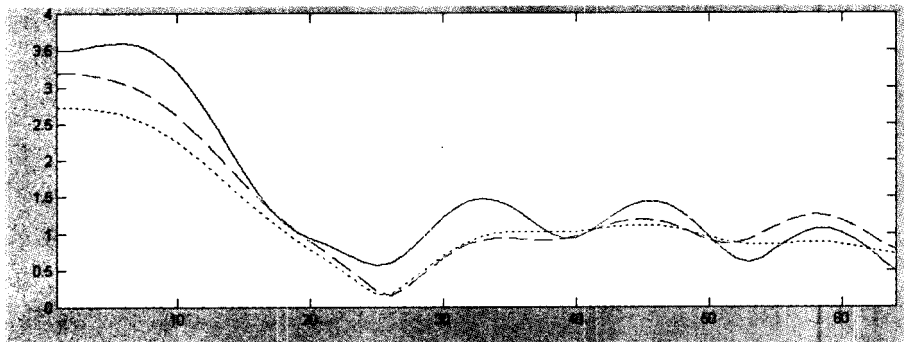
한편 일반적으로 성도 모델로 사용되는 선형 예측 시스템의 선형 예측 계수들을 변환하여 복소 켄스트럼 $\hat{x}[n]$ 을 얻을 수 있다.

$$\hat{x}[n] = a_n + \sum_{k=1}^n \left(\frac{k}{n}\right) \hat{x}[k] a_{n-k}, \quad 1 \leq n \quad (10)$$

그리고 복소 켄스트럼을 푸리에 변환하여 복소 로그리듬 $\hat{X}(e^{j\omega})$ 을 얻을 수 있으며 그 변환식과 정의는 다음과 같다.



(a) 음성의 주파수 응답 (실선)과 α -방법에 의해 극점필터링된 음성 (dash), γ -방법에 의해 극점 필터링된 음성의 주파수 응답 (dotted)
 (a) Comparison frequency responses of speech (solid) and the pole-filtered speech using α -method (dash), γ -method (dotted)



(b) (a) 경우와 동일한 각각의 로그 스펙트럼 형태
 (b) Comparison log magnitude spectra of speech for same cases in (a)

그림 5. 로그 스펙트럼 영역에서의 스무딩된 음성
 Fig. 5. Smoothed speech spectrum in log spectrum domain.

$$\hat{X}(e^{j\omega}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{x}[n] e^{j\omega n} d\omega \quad (11)$$

$$\hat{X}(e^{j\omega}) = \log|X(e^{j\omega})| + j \arg[X(e^{j\omega})]$$

캡스트럼을 푸리에 변환할 때에 변환시키는 특징벡터로 사용하는 차수의 캡스트럼 계수 (실험에서는 12차를 사용)를 이용하므로 푸리에 변환을 거친 로그 스펙트럼은 매우 스무딩된 형태로 나타나게 된다. 이는 음성발성에서의 여기신호가 제거됨을 의미하며 동시에 성도 모델에만 충실한 스펙트럼을 보여준다고 할 수 있다. 그림 5(a)에 해당하는 각각의 캡스트럼을 푸리에 변환하여 얻어진 로그 스펙트럼을 그림 5(b)에 나타내었다. 이 그림을 통해 첫째, 예측 계수로부터 얻어진 주파수 응답에서의 포먼트의 위치가 스무딩된 형태로 잘 나타나며 둘째, 각각의 극점 필터링 효과도 각 포먼트에서 동일하게 표현됨을 볼 수 있다.

제안된 알고리즘은 우세 극점을 구하는 다항식의 계산을 피하기 위해 이러한 우세 극점 필터링과 포먼트 대역 확장 관계를 이용한다. 제안된 알고리즘은 캡스트럼으로부터 변환된 로그 스펙트럼에서 협대역 포먼트를 선택하는 포먼트 검출과 이를 통해 추정된 근을 이용해서 포먼트 대역확장 캡스트럼을 생성하는 과정으로 구성된다.

3.2. 로그 스펙트럼에서의 포먼트 검출과 근의 추정

식 (11)에 의해 큐프렌시 영역의 캡스트럼은 주파수 영역의 로그 스펙트럼으로 변환된다. 푸리에 변환에 의해 캡스트럼으로부터 얻어진 로그 스펙트럼은 매우 스무딩된 형태가 되므로 피치 정보 없이 포먼트를 쉽게 찾을 수 있다[11]. 따라서 제안된 방법은 우세 극점에 대응되는 선형 예측 다항식의 근을 직접 구하지 않고 로그 스펙트럼에서 포먼트를 검출함으로써 간접적으로 구한다.

그림 6에 로그 스펙트럼 영역에서 포먼트를 검출하는 방법을 나타내었다. 포먼트의 중심주파수, $\tilde{\omega}_k$ 는 스펙트럼의 포락선을 쫓아 미분값의 부호가 변화하는 곳으로 결정된다. 그리고 대역폭, B_k 는 중심주파수의 이득과 약 3 dB 차이가 나는 앞뒤의 주파수들을 찾아 얻는다. 얻어진 중심주파수와 대역폭을 이용해서 선형 예측 다항식의 근을 z_k 라 할 때 추정된 근 \hat{z}_k 를 얻을 수 있다.

$$z_k = |e^{-\pi B_k}| e^{j\omega_k}$$

$$\hat{z}_k = |e^{-\pi B_k}| e^{j\omega_k} \quad (12)$$

이때 포먼트의 대역폭 B_k 가 문턱값 B_{TH} 보다 좁은 협

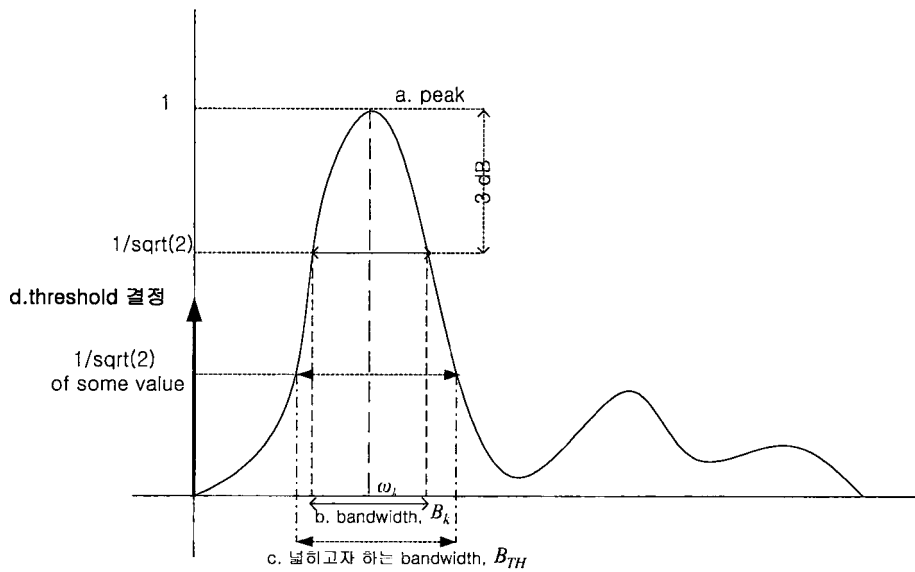


그림 6. 로그 스펙트럼에서 포먼트 대역 확장을 위한 과정
 Fig. 6. The process for formant broadening in log spectrum.

대역 포먼트로 판단될 경우 추정근 \hat{z}_k 를 새로운 근 \hat{z}_k 로 대체하여 대역 확장된 포먼트로 변형시키게 된다. 대체되어야 할 근 \hat{z}_k 는 다음 식에 의해 쉽게 얻을 수 있다.

$$\hat{z}_k = |e^{-\pi B_k T}| e^{j\omega_k} \quad (13)$$

3.3. 포먼트 대역확장 쉰스트럼 변환

문턱값 이하의 협대역 근을 문턱값 대역폭을 가진 새로운 근으로 대체하여 최종 결과인 쉰스트럼을 구하는 과정이다. 이 과정은 다음과 같은 선형 예측 다항식의 근 z_i 와 쉰스트럼 c_n 의 관계를 이용한다[14].

$$\begin{aligned} c_n &= \frac{1}{n} (z_1^n + z_2^n + z_3^n + \dots + z_p^n) \\ &= \frac{1}{n} \sum_{i=1}^p z_i^n, \quad i=1, 2, 3, \dots, p \end{aligned} \quad (14)$$

그러나 제안된 방법은 p 개의 근을 직접 구하는 것이 아니라 협대역 포먼트에 해당되는 근만을 추정하고 새로운 근으로 대체하므로 식 (14)를 다음과 같이 변형하여 협대역 포먼트에 의한 영향만을 평활화한 새로운 쉰스트럼을 구할 수 있다.

$$\begin{aligned} \hat{c}_n &= \frac{1}{n} \{ (z_1^n + z_2^n + z_3^n + \dots + z_p^n) \\ &\quad - (\hat{z}_1^n + \hat{z}_2^n + \dots + \hat{z}_K^n) \\ &\quad + (\hat{z}_1^n + \hat{z}_2^n + \dots + \hat{z}_K^n) \} \end{aligned} \quad (15)$$

이때 협대역 포먼트로서 추정된 K 개의 근은 \hat{z}_k 이고, 대역확장 된 후의 근은 \hat{z}_k 이다. 다시 대체해야 할 근만 고려한 새로운 쉰스트럼 계수는

$$\hat{c}_n = c_n + \frac{1}{n} \left\{ \sum_{k=1}^K \hat{z}_k^n - \sum_{k=1}^K \hat{z}_k^n \right\} \quad (16)$$

와 같이 간략히 표현할 수 있으며 이때 $\hat{z}_k = |e^{-\pi B_k}| e^{j\omega_k}$, $\hat{z}_k = |e^{-\pi B_k}| e^{j\omega_k}$ 이다.

만약 실제 협대역 포먼트의 근 z_k 와 추정된 협대역 포먼트의 근 \hat{z}_k 가 동일하다면 선형 예측 시스템의 근을 직접 문턱값 처리하여 식 (14)에 의해 얻은 것과 동일한 쉰스트럼 계수를 식 (15), (16)을 통해서 얻을 수 있다. 이상의 과정을 통해 얻은 포먼트 평활화된 쉰스트럼을 평균하여 이를 최초의 선형 예측 계수에 의한 쉰스트럼에서 차감하여 채널의 영향을 줄인 쉰스트럼을 얻을 수 있다. 이를 정리하면 다음과 같다.

- i. 쉰스트럼을 푸리에 변환하여 로그 스펙트럼 영역으로 전환한다.
- ii. 로그 스펙트럼 영역에서 포먼트를 찾는다. (중심주파수는 ω_k , 대역폭은 B_k)
- iii. 중심주파수와 대역폭에 의해 근사한 근 \hat{z}_k 를 얻고 B_{TH} 이하의 B_k 의 포먼트인 경우 새로운 근 \hat{z}_k 를 정의한다. (식 (13))
- iv. 기존의 근과 대역 확장된 새로운 근을 통해서 새로운 대역 확장 쉰스트럼 계수를 얻는다. (식 (16))
- v. 대역확장 쉰스트럼 계수를 평균하여 채널 성분 쉰스트럼을 얻고 이를 차감한다.

그림 7에서는 쉰스트럼으로부터 로그 스펙트럼 영역으로 변환하여 근을 계산한 후 포먼트 평활화 쉰스트럼 계수 (FBCC)를 얻는 과정을 보여준다. 그림 4(a)와 비교하였을

때 α -방법에서 선형예측 계수로부터 다항식을 풀어 근을 구하는 부분이 로그 스펙트럼에서 포먼트를 찾는 것으로 대체된다. 선형 예측 다항식으로부터 극점을 필터링하기 위해선 선형 예측 계수의 차수만큼의 근을 모두 구한 후 일정 문턱값을 넘는 값을 찾아내어야 한다. 이에 비해 제안된 방법은 문턱값 이하의 대역폭을 가진 포먼트만을 찾아 식 (12)와 (13)에 의해서 성도 모델의 근으로 변환한다.

실제 구현에서는 로그 스펙트럼의 주파수 정밀도는 변환을 위해 사용하는 고속 푸리에 변화 (FFT)의 크기에 의존하며 높은 정밀도를 얻기 위해서는 많은 연산량을 감수해야 하기 때문에 실제 근과 추정된 근은 오차를 갖게 된다. 그러나 정확한 근을 얻는 것과 채널 정규화의 성능이 반드시 일치하는 것이 아니므로 실제 근과 추정근의 오차

는 무시될 수 있다. 한편 선형 예측 계수의 차수가 p 일 때 최대 근의 수는 $(\frac{p}{2} + 1)$ 이며 이는 실근은 최대 2개일 수 있고 나머지 복소근은 켈레근을 고려해야 하기 때문이다. 제안된 방법은 포먼트를 중심으로 처리되므로 중심 주파수에 의해 처리될 포먼트를 선택하게 되는데 유성음 일 경우 3~4개의 포먼트를 선택하고 포먼트의 중심주파수의 분포 등을 고려하여 유성음 이외의 구간에서는 협대역의 포먼트라도 무시할 수 있다.

그림 8은 TIMIT 데이터베이스의 한 화자에 대한 처리 결과를 보여준다. 한 프레임에 쉼스펙트럼에 대해서 각각 LPCC, PFCC- α , PFCC- γ , 그리고 FBCC 알고리즘에 의해 처리한 후 이를 로그 스펙트럼으로 변환하여 나타내었다.

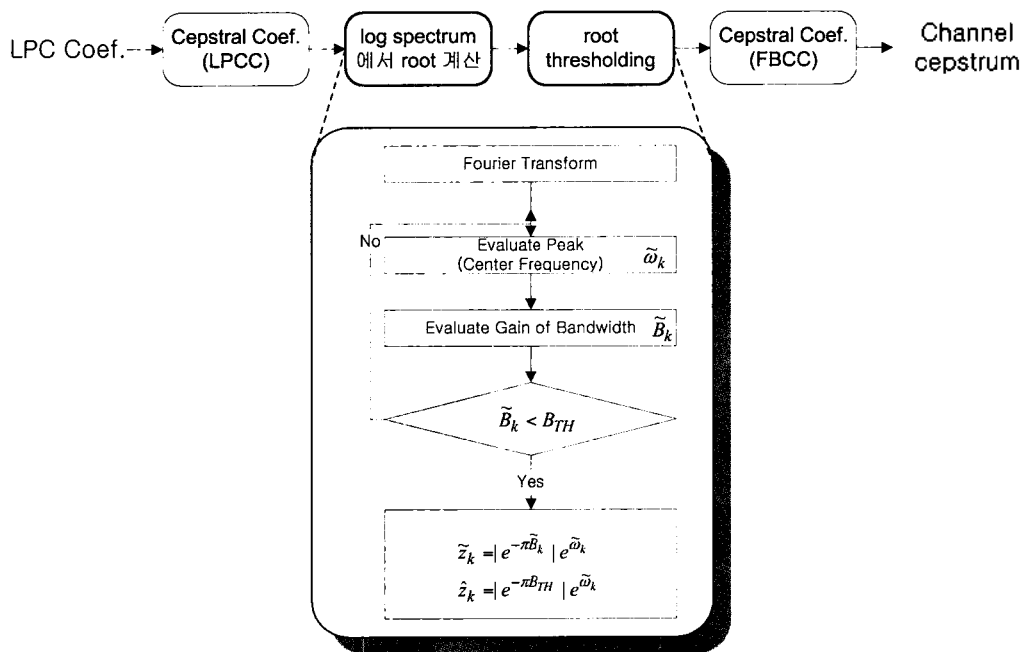


그림 7. 제안하는 방법에서 채널추정 순서도
Fig. 7. Channel estimation in proposed method.

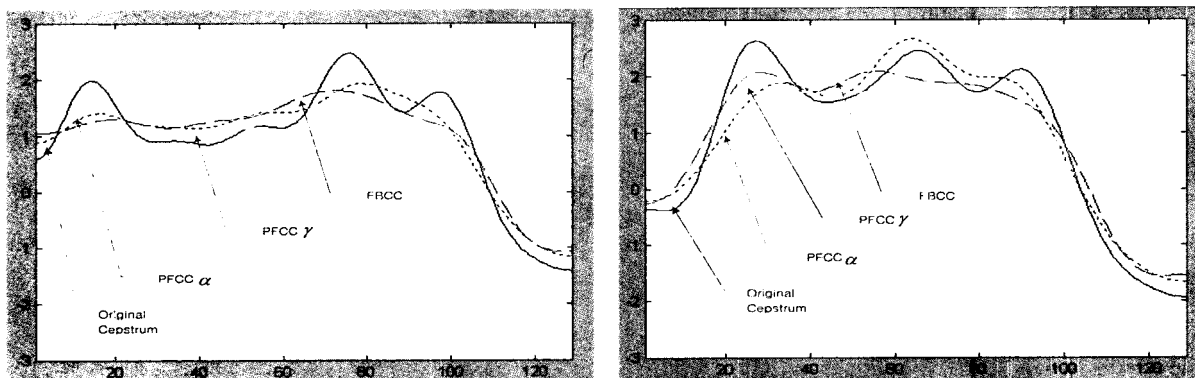


그림 8. LPCC, PFCC- α , PFCC- γ 그리고 FBCC의 로그 스펙트럼
Fig. 8. The log-spectra of LPCC, PFCC- α , PFCC- γ and FBCC.

IV. 실험 결과

제안된 방법을 평가하기 위해 TIMIT 데이터 베이스를 이용하여 기존의 CMS, PFCMS, 그리고 포먼트 평활화 킵스트림 평균 차감법 (FBCMS)의 성능을 비교하였다. 비교를 위한 구체적인 방법에 있어서 첫 번째로 각각의 방법에 의해 추정된 채널이 얼마나 실제 채널에 근사한지를 분석하기 위해 킵스트림 거리를 구하였다. 이는 실제 인식 실험 이전에 채널 추정에 대한 비교를 정량적으로 수행하기 위한 것이며 인식 결과와 반드시 일치한다고 보기는 어렵다. 두 번째 평가 방법으로서 38명에 대한 문장 독립 화자 인식실험을 수행하였다. 마지막으로 제안된 방법과 기존 방법의 연산량 비교를 위해 제안된 방법과 PFCMS- α 방법을 MATLAB으로 구현하고 비교하였다.

4.1. 음성 데이터 베이스 및 채널 시뮬레이션

음성 데이터는 기존 PFCMS와의 성능 비교를 위해 Naik가 사용한 TIMIT DR1 영역의 훈련용 음성 데이터들을 선택하여 사용하였다. DR1 영역의 데이터들은 38명의 화자가 동일한 2 문장 (SA)과 각기 다른 8문장 (SI, SX) 등 총 10문장을 깨끗한 환경에서 발성한 것들이다. 각 문장들은 음향학적인 분포를 적절히 고려하여 작성된 것들이다[6].

순수하게 채널의 영향만을 받은 데이터를 얻기 위하여 음성 데이터를 전화선 채널 시뮬레이터에 통과시켰다. 전화선 시뮬레이터는 대륙성 중음성 (CMV: continental mid Voice), 대륙성 저음성 (CPV: continental poor voice), 유럽식 중음성 (EMV: european mid voice), 유럽식 저음성 (EPV: european poor voice) 등 모두 4개로 통과 대역폭, 통과 대역 감쇠, 전이구간 길이, 전이구간 감쇠 등에

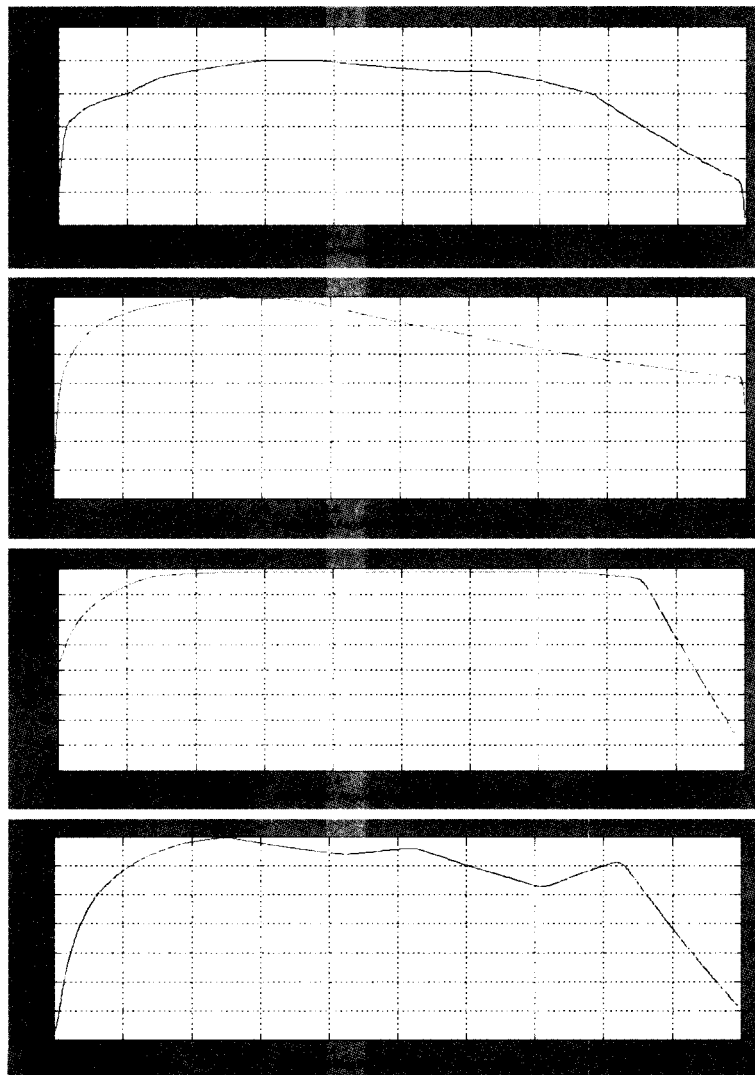


그림 9. 각 채널의 주파수 응답 (왼쪽 상단으로부터 시계 방향으로 CMV, CPV, EMV 그리고 EPV)

Fig. 9. The frequency responses of channels (The clockwise direction from the top of left: CMV, CPV, EMV and EPV).

따른 특성을 가지고 있다[15].

실제 실험에서는 16 KHz로 샘플링된 음성 데이터를 8 KHz로 다운 샘플링하고 다시 201개 계수로 구성된 FIR 필터 형식의 채널 시뮬레이터를 거쳤다. 그리고 채널을 통과한 음성 데이터로부터 20 ms 길이에 대해 10 msec 간격으로 선형 예측 분석을 수행하여 12차의 LPC 특징벡터를 추출하였다. PFCMS α -방법을 구현하기 위해 복소근을 얻을 수 있는 고유값 알고리즘을 사용하여 선형 예측 계수로 이루어진 다항식의 근을 얻었다. 또한 FBCMS 방법을 구현하기 위해 256 크기의 고속 푸리에 변환을 이용하여 로그 스펙트럼을 얻었다.

4.2. 캡스트럴 거리

본 실험은 채널 통과 음성의 캡스트럼으로부터 추정된 채널 성분이 실제 채널과 얼마나 가까운지를 알아보기 위해 수행되었다. 캡스트럴 거리는 각 채널에 대하여 실제 채널의 캡스트럼, c_{ch} 과 추정된 채널의 캡스트럼, 즉 채널 통과 음성의 캡스트럼 평균, \overline{c}_y 사이의 거리로서 다음과 같은 식 (13)에 의해서 정의된다. 이 캡스트럴 거리는 로그

스펙트럴 거리의 제곱에 대응되는 측정 방법이다[16].

$$d^2 = \sum_{n=1}^P (c_{ch} - \overline{c}_y)^2 \tag{13}$$

본 실험에서는 CMV, CPV, EMV 그리고 EPV 등의 4개의 채널 통과 음성으로부터 CMS, PFCMS- α , PFCMS- γ , 그리고 FBCMS 방법으로 채널을 추정하여 실제 채널과의 거리를 구하였다. 그림 10부터 13까지 각 채널에 대해 화자 38명 발성 전체의 캡스트럴 거리 평균을 문턱값의 변화에 따라 제시하였으며 그림 14에서 4개 채널에 대한 평균을 나타내었다. 전체적으로 FBCMS 방법이 CMS, PFCMS γ 방법보다 실제 채널과 근사하며 PFCMS α 방법보다는 떨어지는 것으로 나타났다. 이는 각 프레임에서의 음성 성분인 우세 극점을 제거하는 방법의 차이 때문이며 방법이 정확할수록 실제 채널과의 유사도가 높음을 보여준다.

4.3. 화자 식별

두 번째 실험은 첫 번째 실험에서 사용한 4개의 채널을 통과시킨 채널 불일치 음성을 이용한 폐쇄집단 문장독립

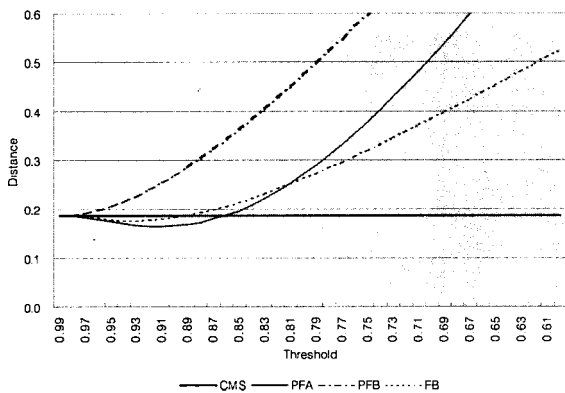


그림 10. CMV 채널의 캡스트럴 거리
Fig. 10. The cepstral distances on CMV channel.

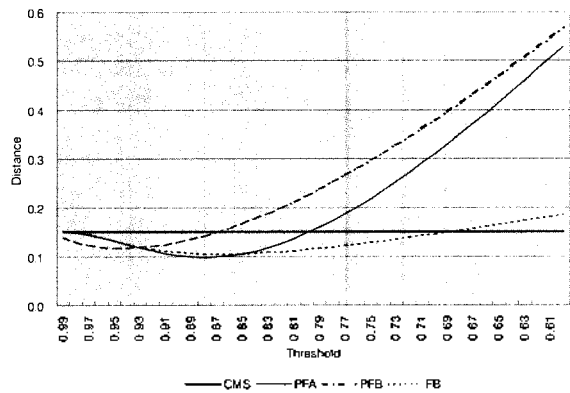


그림 11. CPV 채널의 캡스트럴 거리
Fig. 11. The cepstral distances on CPV channel.

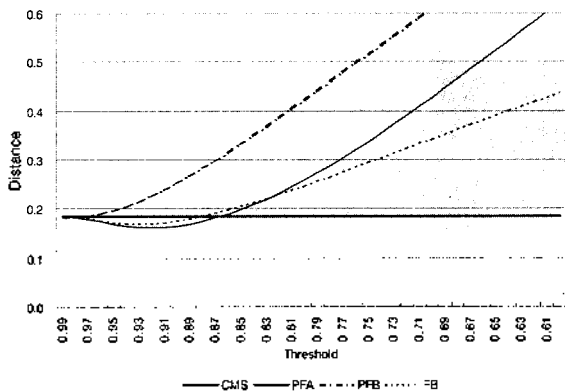


그림 12. EMV 채널의 캡스트럴 거리
Fig. 12. The cepstral distances on EMV channel.

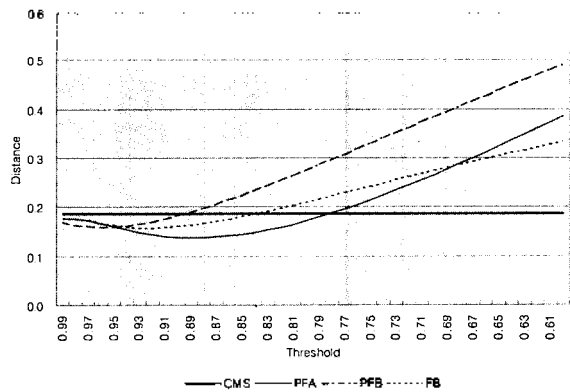


그림 13. EPV 채널의 캡스트럴 거리
Fig. 13. The cepstral distances on EPV channel.

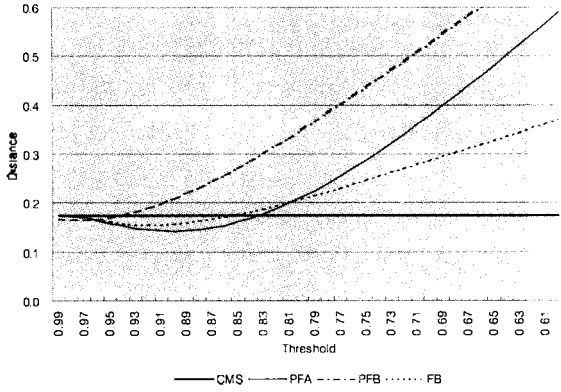


그림 14. 쉰스트럴 거리의 평균
Fig. 14. The average cepstral distances.

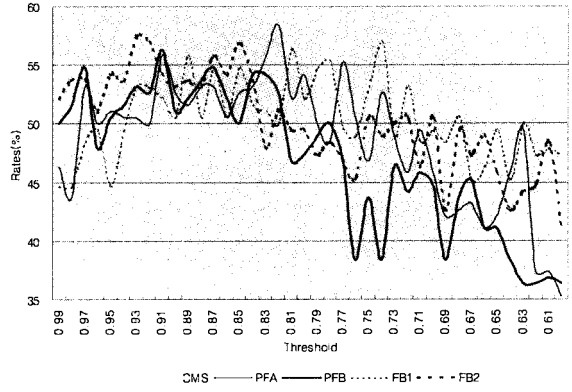


그림 15. CMV (훈련)/CPV (인식) 조건에서의 화자 식별
Fig. 15. The speaker identification in CMV (for training)/CPV (for recognition) condition.

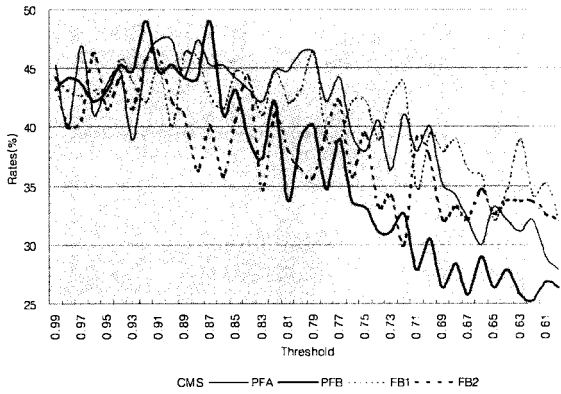


그림 16. CPV/CMV 조건에서의 화자 식별
Fig. 16. The speaker identification in CPV/CMV condition.

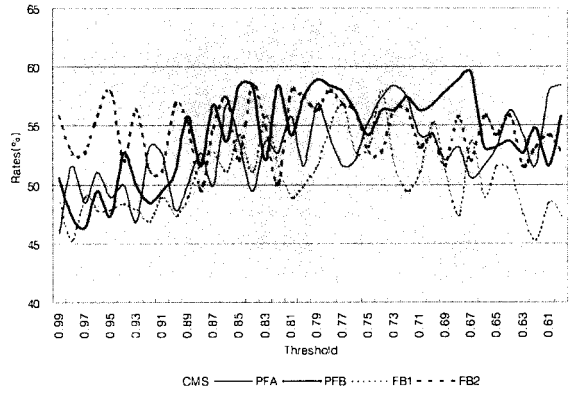


그림 17. EMV/EPV 조건에서의 화자 식별
Fig. 17. The speaker identification in EMV/EPV condition.

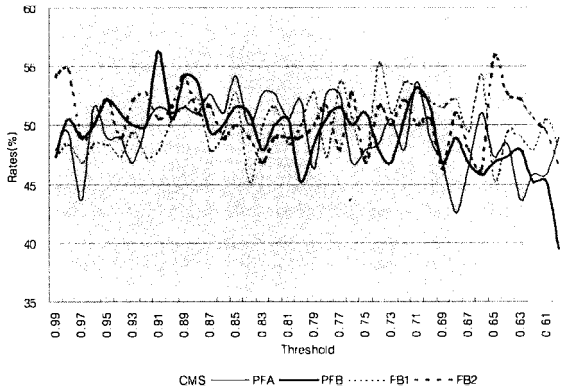


그림 18. EPV/EMV 조건에서의 화자 식별
Fig. 18. The speaker identification in EPV/EMV condition.

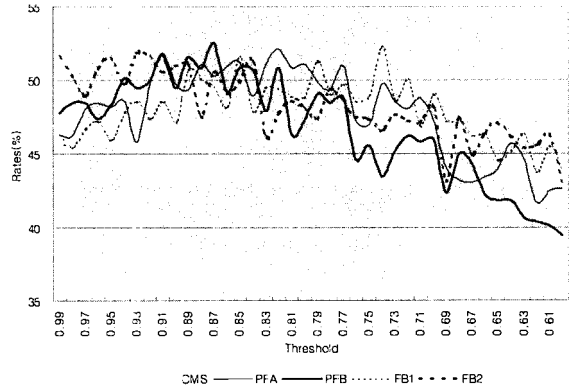


그림 19. 인식률의 평균
Fig. 19. The average identification rates.

화자식별을 통해 실제 불일치 조건에서의 성능을 비교하기 위해 수행되었다. 인식기로는 Naik의 방법과 동일하게 46개의 혼합 (mixture)으로 구성된 가우시안 혼합 모델 (GMM: Gaussian Mixture Model) 알고리즘을 사용하였고 음성 특징으로는 12차의 LPCC만을 사용하였다. 총 38명의 화자 발성 가운데 훈련용으로 SX발성을 사용하였

고 나머지 인식용으로 SI, SA발성을 사용하였다. 그림 15부터 18까지 4가지 채널 불일치 환경에서의 문턱 값 변화에 대한 인식 결과, 그림 19에서 각 문턱값에 대한 모든 결과의 평균을 나타내었다. 전체적으로 PFCMS 방법과 제안된 방법이 모두 기존의 CMS에 비해 뛰어난 인식률을 보여주었다. 하지만 채널 추정의 정확도를 측정할 채널

캡스트럴 거리 실험의 결과와는 일치되는 결과를 보이지 않음을 확인할 수 있다. 즉 PFCMS- α , FBCMS 그리고 PFCMS- γ 방법의 순서로 채널 추정의 정확도를 보였지만 실제 인식 결과에서는 큰 차이를 보이지 않고 최고 인식을 위해서는 오히려 PFCMS- γ 이 가장 높게 나타났다. 이는 채널 추정의 정확도로는 훈련과 인식 환경의 불일치로 인한 오류를 직접적으로 설명할 수 없음을 나타내며 상대적으로 채널 추정이 정확하지 않으며 오히려 불일치를 최소화 하는 것이 인식률에 도움을 주는 것으로 판단된다.

한편 채널에 따른 인식률을 살펴볼 때 CMS의 경우 대륙성에 비해 유립식 채널에서 보다 높은 성능을 보이며 그 외 개선된 CMS 방법의 경우 CMV와 EMV 채널에서 훈련된 경우에 가장 높게 나타났다. 또한 상대적인 차이를 살펴보면 CMV와 EMV에서 가장 커지며 CPV와 EPV에서 보다 작은 차이를 보인다. 특히 CPV/CMV 불일치 환경에서는 기존의 CMS 방법과 개선된 방법이 큰 차이를 보이지 않고 문턱값에 따라 오히려 낮은 인식률을 보였다. 결론적으로 채널이 안정적일수록 기존의 방법에 비해 큰 성능 향상을 보이며 통과 대역에서 가장 큰 감쇄를 보이는 CPV 훈련 환경에서 가장 작은 성능 향상을 보이는 것으로 판단된다.

그림 19의 전체 채널의 평균을 살펴볼 때 PFCMS- γ 가 최고 인식률을 보이지만 전체적으로 문턱값에 따라 편차가 큰 편이며 PFCMS- α , FBCMS는 보다 안정된 성능을 보이는 것으로 나타났다. 하지만 제안된 FBCMS 방법은 전체 평균에서는 기존 방법보다 낮은 성능을 보이는 것으로 나타났으며 현재의 실험 환경에서는 채널 환경 불일치에 대한 포먼트 중심의 정규화가 기존 all-pole 모델의 극점 중심의 정규화에 비해 뛰어나지 못한 것으로 사료된다. 따라서 기존의 포먼트 평활화 방법 (FB1)과 PFCMS- γ 와 결합한 형태의 방법 (FB2)에 대한 실험을 시도하였는데, 이는 PFCMS- γ 의 대역 확장 정규화 방법이 PFCMS- α 의 협대역 극점 필터링보다 전체적으로 뛰어난 성능을 나타낸 결과에 근거한 것이며 결과적으로 제안된 방법의 성능을 어느 정도 향상시킬 수 있는 것으로 나타났다.

결론적으로 우세 극점을 보다 정확하게 필터링함으로써 캡스트럴 거리에서 가장 우수했던 PFCMS- α 과 포먼트 중심의 제안된 정규화 방법은 오히려 가장 단순한 극점 대역 확장 방법인 PFCMS- γ 보다 뛰어난 성능을 보이지 않았다. 이는 실제 채널과 가장 가까운 채널 성분을 구하는 것이 불일치 조건을 정규화해 줄 만큼 뛰어나지 않은 것을 의미하며 보다 적응적인 알고리즘이 필요할 것으로 사료된다.

4.4. 연산량

마지막 실험은 성능과 비례하여 연산량의 정도를 알아보기 위해 수행되었으며 PFCMS- α 방법과 FBCMS 방법에 대해서만 연산량을 비교하였는데, 이는 PFCMS- γ 의 경우 상대적으로 PFCMS- α 방법과 FBCMS 방법에 비해 연산량이 매우 작기 때문이다. 이는 앞으로의 성능 향상이 PFCMS- γ 과 비교할 만큼 작은 연산량을 함께 갖추어야 함을 의미한다고 할 수 있다.

우선 수학적 연산량을 계산해 보면 PFCMS- α 방법에서 사용한 고유값 알고리즘의 알려진 연산량은 $O(3n^3)$ 이며, 이때의 n 은 LPC 계수의 차수이다. 그리고 FBCMS 방법에서 근을 구하는 연산량은 고속 푸리에 변환 연산과 로그 스펙트럼의 피크를 찾는 연산이며, 연산량의 대부분을 차지하는 고속 푸리에 변환 연산량은 잘 알려진 대로 $O(N \log_2 N)$ 이며 이때의 N 은 고속 푸리에 변환의 크기이다. 12차 LPCC를 사용하고 256 크기의 고속 푸리에 변환을 사용할 경우 각각 5184, 2048의 연산량 정도를 얻을 수 있으며 이는 FBCMS의 연산량은 PFCMS- α 연산량의 약 40%만으로 가능한 것을 알 수 있다.

실제적인 연산량의 차이를 알아보기 위해 MATLAB을 이용하여 두가지 알고리즘을 구현하고 수행하였다. LPC 알고리즘과 고유값 알고리즘은 MATLAB의 라이브러리를 이용하고, 12차의 LPCC와 256 고속 푸리에 변환을 사용하였을 때 FBCMS의 연산량은 PFCMS- α 연산량의 약 62.5%에 해당하는 것으로 나타났다.

V. 결론 및 앞으로의 연구 방향

본 논문에서는 음성인식이나 화자인식에서의 채널 불일치에 의한 성능저하를 최소화하기 위하여 CMS를 기반으로 효과적으로 채널성분을 추정하고 제거할 수 있는 채널 정규화 방법을 제안하였다. 제안된 FBCMS 방법은 선형 예측 다항식의 우세 극점에 해당하는 근을 구하기 위해 캡스트럼으로부터 변환된 로그 스펙트럼에서의 포먼트를 찾고 대역폭을 확장한 캡스트럼을 얻는 과정으로 이루어진다.

전통적인 CMS, Naik의 PFCMS와의 비교 실험을 수행한 결과, 제안된 방법은 전통적인 CMS에 비해 음성 성분의 영향을 완화시켜 실제 채널에 보다 가까운 캡스트럼을 얻을 수 있음을 보여주었다. 문장 독립 화자 식별에서도 기존의 CMS 방법에 비해 우세한 결과를 나타내었다. 그

1) FLOPS (floating-point operations per second)

리나 기존의 극점 중심의 개선된 CMS 방법인 PFCMS- α 와 PFCMS- γ 방법에 비해서는 두드러진 성능 향상을 나타내지 못했으며 특히 PFCMS- γ 과 비교하여 많은 연산량에 비해서 오히려 낮은 성능을 나타내는 근본적인 단점을 가지는 것으로 나타났다. 이는 훈련과 인식 불일치 환경의 다양한 원인에 대한 포먼트 중심의 정규화가 정확하지 못하며 일관되지 않은 알고리즘의 한계로 사료된다. 최종적으로 제안된 방법은 연산량 측면에서 거의 부담이 없는 PFCMS- γ 방법과의 결합을 통해서 보다 향상된 성능을 얻을 수 있었다.

앞으로 이러한 문제점을 개선하기 위해 우선 각 발생 채널에 따른 적응적 문턱값 결정에 대한 연구를 진행하고 프레임 내에서 뿐만 아니라 프레임 상호간에 처리할 수 있는 기법을 추가할 예정이다. 또한 제안된 방법과 기존 방법의 실시간 정규화 성능에 대한 실험을 진행할 예정이다.

감사의 글

본 연구는 정보통신진흥원의 2001년도 대학기초사업 지원에 의하여 수행되었습니다.

참고 문헌

1. R. J. Mammone, X. Zhang and R. P. Ramachandran, "Robust speaker recognition", *IEEE signal processing magazine*, 58-71, 1996.
2. 조태현, 김유진, 이재영, 정재호, "전화선 채널이 화자확인 시스템의 성능에 미치는 영향", *한국음향학회지*, 18 (5), 7, 1999.
3. 전원석, "음성 인식을 위한 전화망에서의 잡음처리", *한국음향학회지*, 17 (1), 4-8, 1997.
4. A. A. Garcia and R. J. Mammone, "Channel-robust speaker identification using modified-mean cepstral mean normalization with frequency warping", *Proc. ICASSP*, 325-328, 1999.
5. D. Naik, "Pole-filtered cepstral mean subtraction", *Proc. ICASSP*, 157-160, 1995.
6. D. Naik and R. Mammone, "Channel normalization using pole-filtered cepstral mean subtraction", *Proc. SPIE*, vol. 2277, 99-111, July 1994.

7. R. Balchandran, V. Ramanujam, R. J. Mammone, "Channel estimation and normalization by coherent spectral averaging for robust speaker verification", *Proc. Eurospeech*, 755-758, 1999.
8. T. F. Lo, M. W. Mak and K. K. Yiu, "A new cepstrum-based channel compensation method for speaker verification", *Proc. Eurospeech*, 775-778, 1999.
9. Jae H. Chung, *A New Homomorphic Vocoder Framework Using Analysis-by-Synthesis Excitation Analysis*, Ph. D Thesis, School of Electrical Engineering, Georgia Institute of Technology, 1991.
10. L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall Inc., 1978.
11. S. Furui, "Cepstral Analysis Technique for Automatic Speaker Verification," *IEEE ASSP*, ASSP-29, 254-272, April 1981.
12. B. H. Juang, L. R. Rabiner and J. G. Wilpon, "On the use of bandpass filtering in speech recognition," *IEEE ASSP*, ASSP-35, 947-954, July 1987.
13. R. P. Ramachandran and K. R. Farrell, "Fast pole-filtering for speaker recognition," *Proc. ISCAS*, V-49-52, 2000.
14. M. Schroeder, "Direct(nonrecursive) relations between cepstrum and predictor coefficients," *IEEE ASSP*, ASSP-29, 297-301, April 1981.
15. J. Kupin, "A Wireline Simulator[Software]," CCR-P, April 1993.
16. L. R. Rabiner, B. H. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall Inc., 1993.

저자 약력

● 정 해 경 (Hea-Kyoung Jung)



1999년 2월: 인하대학교 전자공학과 (공학사)
 2001년 2월: 인하대학교 전자공학과 (공학석사)
 2001년 2월 - 현재: 국방과학연구소 연구원
 ※ 주관심분야: 음성신호처리, DSP 구현

● 김 유 진 (Yu-Jin Kim)



1995년 2월: 인하대학교 전자공학과 (공학사)
 1997년 2월: 인하대학교 전자공학과 (공학석사)
 1995년 6월 - 1996년 12월: 한국전자통신연구소 음성언어처리연구실 위촉 연구원
 1997년 2월 - 1998년 5월: LG반도체 System Device 연구소 DSP 그룹 연구원
 1998년 9월 - 현재: 인하대학교 전자공학과 박사과정 (수로)

※ 주관심분야: 패턴인식, 음성인식, 화자인식

● 정 재 호 (Jae-Ho Chung)

한국음향학회지 제21권 제2호 참고