Spectral Folding방법과 GMM 변환을 이용한 대역폭 확장의 Hybrid 방법

최무열, 김형순
부산대학교 전자공학과

The Hybrid Bandwidth Extenstion Method Using Spectral Folding and GMM Transformation

Mu Yeol Choi, Hyung Soon Kim
Dept. of Electronics Engineering, Pusan National University
{mychois, kimhs}@pusan.ac.kr

Abstract

The narrowband speech over the telephone network is lacking in the information from low-band (0-300 Hz) and high-band (3400-8000 Hz) that are found in wideband speech (0-8000 Hz). As a result, narrowband speech is characterized by the reduced intelligibility and muffled quality, and degraded speaker identification. Spectral folding is the easiest way to reconstruct the missing high-band; however, the reconstructed speech still brings the sense of band-limited characteristic because of the absence of low-band and mid-band frequency components. To compensate for the lack of the extended speech, we propose to combine the spectral folding method and GMM transformation method, which is a statistical method to reconstruct wideband speech. The reconstructed wideband speech showed that the absent frequency components was filled up with relatively low spectral mismatch. According to the subjective speech quality evaluations, the proposed method was preferred to other methods.

Ⅰ. 서론

아날로그 전화망과 이동 통신망을 포함해 현존하는 대부분의 음성통신 시스템은 300-3400 Hz 대역의 협대역(narrowband) 음성신호를 전송한다. 이러한 협대역 음성은 0-8000 Hz의 광대역 신호와 비교할 때 0-300Hz의 저대역과 3400-8000 Hz의 고대역 성분이 제거된 특성으로 인해 명료도가 감소되고 억눌린(muffled) 음질을 갖는다. 대역폭 확장(bandwidth extension)은 협대역 음성신호의 음질을 향상시키는 기술로서 제거된 저대역과 고대역의 음성 신호를 추정하여 복원함으로써 대역폭을 확장한다. 협대역 음성으로부터 광대역 음성을 복원하는 방법에 대한 연구들은 음성부호화기로 선형예측(LPC)방법을 사용할 경우 스펙트럼 포락선의 추정[1][2][3]과 여기신호의 발생[4][5][6]으로 나눈다.

LPC를 음성부호화기로 사용하지 않는 방법 중에서 spectral folding방법은 대역폭을 확장하는 가장 간단한 방법으로 스펙트럼 포락선과 여기신호의 대역폭 확장에 사용되어 왔다. 그러나 이 방법은 300-3400 Hz의 협대역 신호를 사용할 경우 0-300 Hz의 저대역과 3400-4500 Hz의 중간대역의 손실이 여전히 존재하여 0-8000 Hz까지의 완전한 광대역 음질을 복원하지 못하는 문제가 있다.

본 논문에서는 spectral folding을 이용한 대역폭 확장 방식[7]의 개선을 위해 이 방법과 GMM 변환 방법의 hybrid를 통한 대역폭 확장을 제안한다. 본 논문의 구성은 다음과 같다. 2장에서 기존의 GMM을 이용한 대역폭 확장 방법에 대해 설명하고, 3장에서 cubic spline 보간을 적용한 spectral folding 방법에 대해 설명한다. 4장에서는 두 가지 방법의 hybrid 방법을 제안하고, 5장에서 실험결과를 보이며, 6장에서 결론을 맺는다.

II. GMM을 이용한 대역폭 확장

협대역 신호로부터 광대역 신호를 추정하는 GMM방법은 협대역 신호와 광대역 신호가 서로 상관관계에 있다는 가정 하에 출발한다.

협대역 신호를 $\boldsymbol{x} \in R^n$이라 하고 추정할 광대역 신호를

$\boldsymbol{y} \in R^n$이라고 하면 $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})^T$는 Q 개의 Gaussian 확률밀도함수를 이용한 GMM으로 모델링 된다.

$$p(\boldsymbol{z} \mid \boldsymbol{\lambda}) = \sum_{i=1}^{Q} \frac{\alpha_i}{(2\pi)^{n/2} \left| \boldsymbol{C_i} \right|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} (\boldsymbol{z} - \boldsymbol{\mu_i})^T \boldsymbol{C_i}^{-1} (\boldsymbol{z} - \boldsymbol{\mu_i}) \right]$$

$$\sum_{i=1}^{Q} \alpha_i = 1, \ \alpha_i \geq 0 \qquad (1)$$

여기서 α_i, $\boldsymbol{\mu_i}$ 그리고 $\boldsymbol{C_i}$는 i번째 밀도함수의 가중치, 평균벡터, 그리고 공분산행렬을 나타내며, 다음 식과 같이 자승오차를 최소화하는 mapping 함수를 통해 광대역 스펙트럼 포락선을 추정한다.

$$\epsilon_{mse} = E[\| \boldsymbol{y} - F(\boldsymbol{x}) \|^2] \qquad (2)$$

여기서 $E[.]$는 기대값을 나타내며, $F(\boldsymbol{x})$는 추정될 광대역 스펙트럼 포락선이 된다.

최소자승오차를 만족하는 mapping 함수는 다음과 같이 표현된다.

$$F(\boldsymbol{x}) = E[\boldsymbol{y} \mid \boldsymbol{x}] = \sum_{i=1}^{Q} h_i(\boldsymbol{x}) \boldsymbol{\mu_i^y} \qquad (3)$$

여기서

$$h_i(\boldsymbol{x}) = \frac{\frac{\alpha_i}{(2\pi)^{n/2} \left| \boldsymbol{C_i^{xx}} \right|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu_i^x})^T \boldsymbol{C_i^{xx^{-1}}} (\boldsymbol{x} - \boldsymbol{\mu_i^x}) \right]}{\sum_{i=1}^{Q} \frac{\alpha_i}{(2\pi)^{n/2} \left| \boldsymbol{C_i^{xx}} \right|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu_i^x})^T \boldsymbol{C_i^{xx^{-1}}} (\boldsymbol{x} - \boldsymbol{\mu_i^x}) \right]}$$

및 $\boldsymbol{C_i} = \begin{bmatrix} \boldsymbol{C_i^{xx}} & \boldsymbol{C_i^{xy}} \\ \boldsymbol{C_i^{yx}} & \boldsymbol{C_i^{yy}} \end{bmatrix}$, $\boldsymbol{\mu_i} = \begin{bmatrix} \boldsymbol{\mu_i^x} \\ \boldsymbol{\mu_i^y} \end{bmatrix}$ (4)

이며, $h_i(\boldsymbol{x})$ 는 i번째 Gaussian 밀도 함수의 사후 확률을 나타낸다.

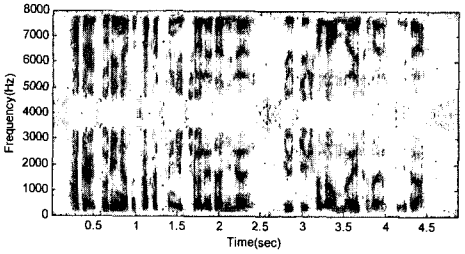
LPC 모델에 의한 음성합성은 합성 필터와 여기신호를 통하여 음성을 합성해 낸다. 본 논문에서는 고대역의 주기성과 비주기성을 통계적인 방법을 통해 추정하고 임펄스와 잡음 성분에 가변적인 비율을 적용하는 방식을 사용하여 여기신호를 발생시켰다[6].

대역폭 제한된 신호의 에너지는 특히 무성음처럼 고대역에 큰 에너지를 갖는 경우 대부분의 에너지를 잃게 되어 명료성이 저하된다. 본 논문에서는 고대역 및 저대역 복원신호의 에너지를 보다 정교하게 추정하기 위해, 음향공간을 유성음 구간과 무성음 구간으로 나누고, 각각 GMM으로 모델링 하여 입력된 협대역 신호를 통해 저대역과 고대역의 에너지를 추정하였다[6].

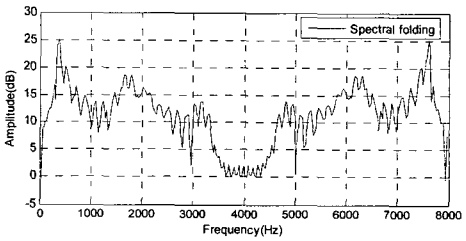
III. Cubic Spline 보간을 이용한 Spectral Folding 방법

음성 신호의 대역폭을 확장시키는 가장 간단한 방법은 샘플링 rate을 변경하는 방법으로 전화음성 신호의 샘플 rate을 2배로 증가 시키면 narrowband의 주파수 영역의 정보가 highband 영역에 나타나게 된다. 이로 인해 음성신호의 대역폭이 확장되는데 이와 같은 현상을 spectral folding이라고 부른다[4]. 그러나 이 방법에는 다음의 두 가지 문제점이 존재한다.

첫째로, 전화음성 대역의 spectral folding은 그림 1의 (a), (b)에서 보는 유성음의 예와 같이 협대역의 하모닉 성분이 고대역에 동일하게 나타나므로 하모닉의 크기와 주기성에 의해 고대역의 음질을 저하시킨다. 둘째로, spectral folding을 통해 고대역의 스펙트럼 성분을 어느 정도 복원한다 하더라도 그림 1의 (a), (b)에서 보는 바와 같이 0-300 Hz의 저대역과 3400-4500 Hz의 중대역의 손실된 스펙트럼 성분은 여전히 문제점으로 남게 된다.



(a) Spectral folded spectrogram



(b) Spectral folded spectrum

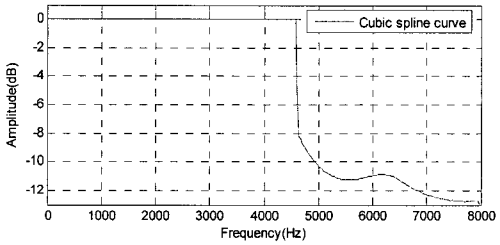
그림 1. Spectral folding에 의한 전화음성의 스펙트로그램과 스펙트럼

첫번째 문제를 해결하기 위해 고대역의 스펙트럼 영역에 cubic spline 보간 방법을 적용하여 음질을 개선하는 방법이 제안되었다[7].

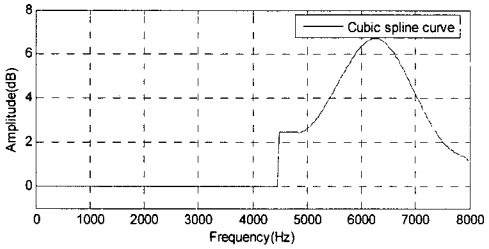
고대역에 나타나는 협대역의 스펙트럼 포락선은 그림 2에서 보는 바와 같이 미리 훈련된 데이터로부터 추정되는 값에 의해 감쇄 또는 증가된다[7]. 본 논문에서는 4-8 kHz 사이에 균등히 위치한 5개의 미리 정해진 점들을 두고 GMM에 의해 추정된 음소 그룹에 따라 대푯값을 두었다. 이 5개의 점들은 cubic spline 보간을 통하여 고대역의 스펙트럼 포락선을 추정하는데 사용되었다.

고대역의 스펙트럼 포락선은 음소에 따라 매우 다양하게 나타난다. 본 논문에서는 이를 유성음과 무성음으로 나누고 유성음은 각각 2개의 그룹, 그리고 무성음

은 파열음과 마찰음에 따라 각각 3개 그룹으로 나누었다. GMM 추정 방법을 통해 구별된 음소 그룹에 따라 적용한 유성음과 무성음 cubic spline 곡선의 예를 그림 2에 나타내었다.



(a) 유성음



(b)무성음

그림 2. 유성음과 무성음에 적용되는 cubic spline 곡선의 예

두번째 문제를 극복하기 위해 본 논문에서는 기존의 GMM을 통한 대역폭 확장방법을 이용하여 스펙트럼 영역의 불연속성을 최소로 하는 스펙트럼 보상 방법을 제안하였다.

IV. Spectral Folding방법과 GMM변환 방법의 Hybrid 방식

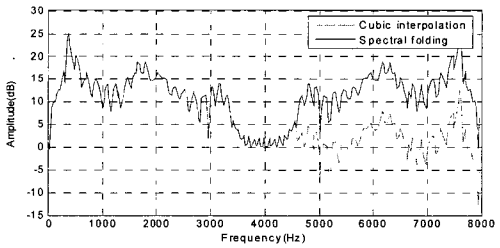
본 논문에서 제안하는 hybrid 방식은 spectral folding으로 복원하지 못하는 0-300 Hz와 3400-4500 Hz 대역의 신호를 0-8000 Hz까지 복원하는 GMM 변환 결과로부터 가져오는 것이다. GMM 변환 방법은 식 (2)에서처럼 복원하는 신호와 원신호의 오차가 최소가 되도록 합성하기 때문에 300Hz와 3400Hz 에서의 스펙트럼 불연속이 감소되는 장점이 있다.

본 논문에서는 GMM 변환 결과에서 저대역과 중간대역의 대역필터를 이용하여 spectral folding결과와 결합하는 방법을 취하였다.

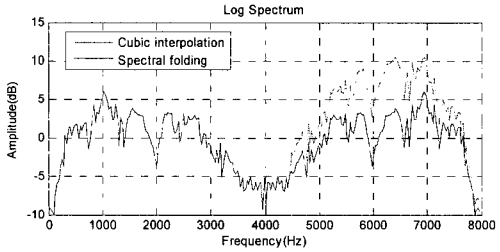
V. 실험 결과

제안된 방법의 실험을 위해 국어공학센터에서 구축한 PBS 데이터베이스중에서 남녀 20명분의 발성 데이터 각각 10문장씩을 사용하였다. 광대역 스펙트럼 변환, 에너지 추정, 여기신호 가중치 추정 그리고 cubic spline 보간 방법을 적용하기 위한 음소 그룹 구별을 위해서 256 mixture를 갖는 GMM 모델을 훈련하였다.

그림 3에서 보는 바와 같이 고대역에 나타나는 하모닉 성분의 크기는 그림 2와 같은 cubic spline 곡선에 의해 감쇄 또는 증가된 것을 확인할 수 있다.



(a) 유성음



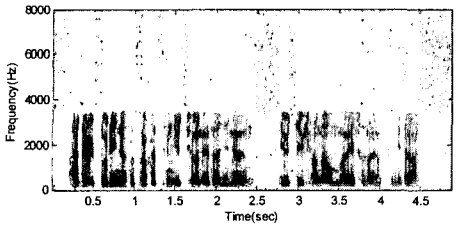
(b) 무성음

그림 3. Cubic spline 곡선에 의해 변형된 유성음과 무성음의 스펙트럼 결과

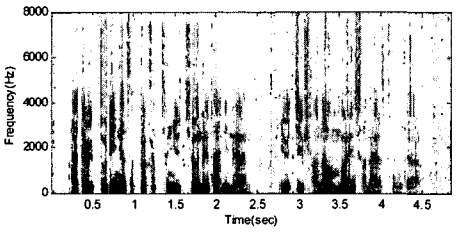
본 논문에서 제안한 hybrid 방법의 대역폭 확장 결과를 그림 4에 나타내었다. 그림 4의 (b)는 GMM을 이용한 대역폭 확장 방법으로 고주파영역의 에너지와 여기신호의 복원 음질에 따라 자연성이 저하되는 경향이 있다. 그림 4의 (c)는 spectral folding 방법에 cubic spline 보간 방법을 적용하여 대역폭이 확장된 음성이고 (d)는 Hybrid 방법의 결과이다. 그림 (d)에서 보는 바와 같이 (c)의 저대역과 중대역의 대역이 보상된 것을 볼 수 있다.

제안된 방법의 음질 평가를 위해 cubic spline 보간 방법을 적용한 spectral folding 방법, GMM 변환 방법, 그리고 제안된 방법의 합성음 중에서 가장 선호하는 음성을 선택하는 주관적 테스트 방법을 사용하였다. 실험에 사용된 음성은 훈련에 사용되지 않은 남성 3명과 여성 3명의 발성 문장을 사용했으며, 세 가지 합성 방식에 따라 각각 6 문장을 합성한 뒤 10명의 청취자에게 헤드셋을 통해 청취하도록 하였다. 그림 5에서 보는 바와 같이 제안한 방식의 결과가 기존 방식들 보다 평균

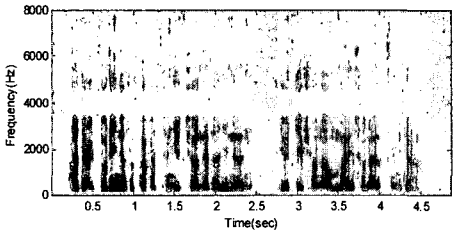
적으로 높은 선호도를 보였다.



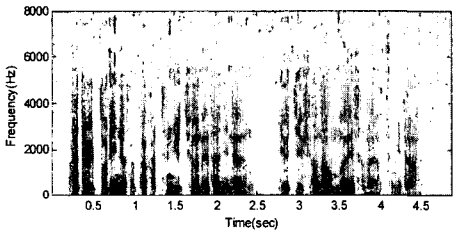
(a) 협대역 음성신호



(b) GMM 변환 방법



(c) Spectral folding 방법



(d) 제안된 방법

그림 4. 협대역 음성 및 세 가지 방법의 대역폭 확장 방법의 스펙트로그램

VI. 결론

본 논문에서는 spectral folding 방법과 GMM 변환 방법의 hybrid 방식에 의한 전화음성의 대역폭 확장 방법을 제안하였다. Spectral folding 방법에 나타나는 저대역과 중대역의 대역 손실이 GMM변환 방법과의 hybrid 방식을 통해 보완되었으며, 기존방식들과 제안된 방식에 대한 주관적인 음질평가에서 제안 방식의 합성음이 평균적으로 높은 선호도를 나타냈다.

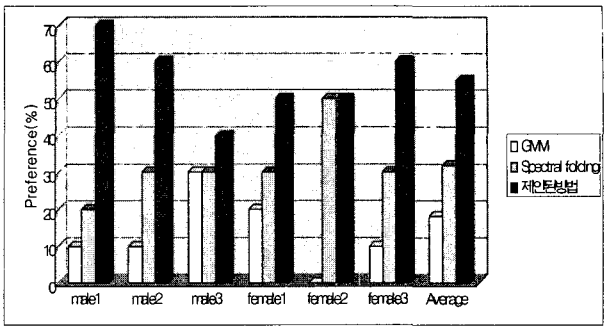


그림 5. 주관적 청취평가 결과

참고문헌

[1] Y. Yoshida and M. Abe, "An algorithm to reconstruct wideband speech from narrowband speech based on codebook mapping," in Proc. of ICSLP, pp. 1591-1594, 1994.

[2] K. Y. Park and H. S. Kim, "Narrowband to wideband conversion of speech using GMM based transformation," in Proc. of ICASSP, vol. 3, pp. 1843-1864, 2000.

[3] P. Jax and P. Vary, "Wideband extension of telephone speech using a hidden markov model," in Proc. of IEEE Workshop on Speech Coding, 2000.

[4] J., Makhoul, M., Berouti, "High-frequency regeneration in speech coding systems," in Proc. of ICASSP, vol. 4, pp. 428 - 431, 1979

[5] Y. Qian and P. Kabal, "Classified Highband Excitation for Bandwidth Extension of Telephony Signals," in Proc. European Signal Processing Conf. 2005.

[6] 최무열, 김형순, "혼합여기모델을 이용한 대역 확장된 음성신호의 음질 개선," 대한음성학회 말소리, 제52호, pp. 133-144, 2004년 12월.

[7] L. Laaksonen, J. Kontio and P. Alku, "Artificial bandwidth expansion method to improve intelligibility and quality of AMR-coded narrowband speech," in Proc. ICASSP, pp. 809-812, 2005.